

**POLITECHNIKA WROCŁAWSKA**  
**WYDZIAŁ PODSTAWOWYCH PROBLEMÓW**  
**TECHNIKI**

**Roksana Wdowiak**

**Identyfikacja struktur sieci złożonych**

**Praca inżynierska**  
**napisana pod kierunkiem**  
**dr Rafała Werona**



## **Spis treści:**

1. Wstęp .....	5
1.1 Wprowadzenie .....	5
1.2 Cel .....	7
1.3 Zakres pracy .....	8
2. Podstawowe pojęcia i definicje .....	10
2.1 Podstawowe pojęcia dotyczące sieci .....	10
2.2 Sieci i ich własności (sieci złożone, struktury sieci) .....	13
2.2.1 Sieć regularna .....	13
2.2.2 Sieć losowa .....	14
2.2.3 Sieci Małego Świata .....	15
2.2.4 Sieci bezskalowe .....	16
2.3 Wykorzystanie badań nad topologią sieci .....	19
2.4 Zastosowanie wiedzy o centrach w sieciach .....	19
2.5 Wykorzystanie badań nad przebiegiem procesów dyfuzji .....	20
2.5.1 Co nazywamy procesem dyfuzji? .....	20
2.5.2 Zastosowanie wiedzy o procesach dyfuzji i ich rozpoznanie w układzie .....	20
2.5.3 Znaczenie badań nad procesami dyfuzji .....	21
3. Technologie użyte do zbudowania programu „Analiza sieci” .....	22
3.1 Matlab jako podstawowe narzędzie programistyczne .....	22
4. Program „Analiza sieci” .....	23
4.1 Formaty danych do programu „Analiza sieci” .....	23
4.2 Opis korzystania z programu .....	24
4.3 Algorytmy stosowane w programie .....	25
4.3.1 Algorytm budowania sieci losowej .....	25
4.3.2 Algorytm budowania sieci zachłannej .....	26
4.3.3 Metodologia tworzenia diagramu obrazującego strukturę sieci pod względem modułowości – badanie procesów dyfuzji .....	26
4.4 Funkcjonalność programu .....	28
4.5 Opis otrzymanych wyników wyjściowych i ich analiza .....	29
4.5.1 Sieć znajomości w sieci Zacharego .....	29
4.5.1.1 Badanie rozkładu .....	31
4.5.1.2 Porównanie sieci wyjściowej ze spreparowaną siecią losową .....	31
4.2.1.3 Porównanie sieci wyjściowej ze spreparowaną siecią, generowaną algorytmem zachłannym .....	32
4.5.2 Sieć znajomości Nasza Klasa 1 .....	33
4.5.2.1 Badanie rozkładu .....	35
4.5.2.2 Porównanie sieci wyjściowej ze spreparowaną siecią losową .....	37
4.5.2.3 Porównanie sieci wyjściowej ze spreparowaną siecią, generowaną algorytmem zachłannym .....	38
4.5.3 Sieć znajomości Nasza Klasa 2 .....	39
4.5.3.1 Badanie rozkładu .....	41
4.5.3.2 Porównanie sieci wyjściowej ze spreparowaną siecią, generowaną algorytmem zachłannym i losowym .....	42
4.5.4 Sieć znajomości Nasza Klasa 3 .....	43

4.5.5 Sieć znajomości i zaznaczone w niej moduły(podgrupy) - Nasza Klasa 4	45
5. Zastosowanie wiedzy o modułach w sieci.....	45
Dodatek.....	49
Zagadnienia dotyczące języka programowania Perl.....	49
Dlaczego język Perl? .....	49
Podstawowa składnia.....	49
Typy danych .....	50
Najważniejsze operatory.....	50
Wyrażenia regularne .....	51
Operacje na plikach .....	51
Matlab jako podstawowe narzędzie programistyczne. ....	52
Podstawowe typy danych w Matlabie .....	52
Definiowanie własnych poleceń i funkcje w Matlabie .....	52
Pomocne makra w Excelu i Microsoft Visual Basic .....	53
Bibliografia:.....	54

# 1. Wstęp

## 1.1 Wprowadzenie

Sieci kryją się wszędzie lub - inaczej - świat jest po prostu siecią. Jako sieć można przedstawić złożone ekosystemy, szlaki metaboliczne, połączenia neuronowe w mózgu, epidemię AIDS, katastrofę energetyczną w Stanach Zjednoczonych w 1996 r., kryzys finansowy lat 1997-1998, walkę o władzę we Florencji epoki Renesansu, strukturę władzy i procesy odsłaniane w Polsce podczas afery Rywina. Co ważniejsze, do opisu tych wszystkich zjawisk sieciowych nadają się podobne modele matematyczne, odkryte m.in. podczas badań nad Internetem. Rzeczywistość ma strukturę sieci i struktura ta wykazuje niezwykle, często kwestionujący powszechne intuicje, porządek. Świat okazuje się często niespodziewanie mały ale przy tym również niezwykle złożony [1].

W otaczającym nas świecie sieci są wszechobecne i przybierają najróżniejsze formy. Ze względu na różnorodność tych tworów nauka nadal jest na etapie poznania tylko ich ogólnych własności. Sprawnie wypracowane narzędzie pozwalające efektywnie symulować i modelować otaczające nas struktury sieciowe mogłoby być źródłem niezliczonej ilości przydatnych informacji w wielu dziedzinach wiedzy. Niestety okazuje się, że modelowanie sieci rzeczywistych przy użyciu wiedzy na temat typowych struktur (sieci regularne, pełne...) nie daje dobrych rezultatów. Stąd nauka skupiła się na badaniu tzw. sieci kompleksowych, których budowa jest dużo bardziej zaawansowana pod względem złożoności ale jednocześnie w dużym stopniu odzwierciedla budowę sieci tak często występujących i obserwowanych w realnym świecie. Własności wielu skomplikowanych pod względem budowy sieci występujących w niezwiązanych ze sobą obszarach nauki i dziedzinach życia są do siebie bardzo podobne. Jednakże tylko dla niektórych sieci złożonych istnieją efektywne algorytmy do ich generowania, chociażby w przybliżeniu oddające rzeczywistą istotę ich budowy.

Cechy sieci zbudowanych kompleksowo i określenie ich struktury a także własności jest procesem trudnym ze względu na obszerną i złożoną budowę tych tworów i co się z tym wiąże potrzebę posiadania ogromnych zasobów wiedzy z dziedziny matematyki, fizyki a

nawet socjologii. Trudności wynikające z próby opisu takich struktur za pomocą wzorów matematycznych oraz duży stopień zaawansowania obliczeń i algorytmów dotyczących budowy dużych sieci o skomplikowanej strukturze pociągają za sobą wymagania niekiedy znacznie przekraczające aktualne możliwości mocy obliczeniowej dobrego komputera.

Badania z ostatnich lat znacznie poszerzyły wiedzę w kierunku sieciowego ujęcia niektórych zjawisk. O sieciach wiedziano już dużo w latach siedemdziesiątych, ale była to głównie wiedza opisowa. Dostępne wówczas modele matematyczne sieci, stworzone m.in. przez jednego z najwybitniejszych matematyków XX w. Paula Erdösa nie wystarczały. Rewolucja nastąpiła pod koniec lat dziewięćdziesiątych, a jej sprawcą w dużej mierze stał się Internet [1].

Dokładniejsze analizy pozwoliły dostrzec, że pomimo coraz większego uzależnienia zachowań ludzi i innych organizmów a także technologii od topologicznych właściwości sieci złożonych do teraz nie w pełni rozumiemy zasady rządzące ewolucją i dynamiką tych układów. Do niedawna jeszcze nauka skupiała się głównie wokół badań na temat sieci regularnych i losowych, podciągając wiele sieci o złożonej budowie pod kategorię tych drugich. Okazało się jednak, że w ogromie sieciowych układów można wyróżnić charakterystyczne grupy struktur, które mogą stanowić odrębne rodzaje sieci. Analogia w algorytmach do budowy tych sieci stała się główną miarą przynależności do konkretnej grupy.

Symulacje przeprowadzone w ostatnich latach dowodzą, że wystarczy dodać kilka przypadkowych połączeń w regularnej sieci, aby znacznie przyspieszyć w niej przepływ informacji. Tak powstała koncepcja opracowania algorytmu budowy sieci „Małego Świata” poprzez zaburzenie struktury sieci regularnej wprowadzając niewielką ilość losowych połączeń. Sieć regularna (np. sieć atomów w kryształach) gdzie każdy element ma stałą liczbę połączeń z sąsiadami, i sieć o połączeniach przypadkowych - to dwa skrajne przykłady sieci pod względem rodzaju połączeń. Pomędzy nimi znajdują się sieci typu „Małego Świata”, z właściwościami sieci przypadkowej i regularnej. Jak się okazuje, na układy tego rodzaju można natknąć się w wielu przypadkach: powiązania między aktorami filmowymi (połączenie bezpośrednie istnieje, jeśli grali w tym samym filmie), sieć elektryczna w USA i sieć neuronalna organizmów żywych [2].

Co jednak możemy powiedzieć o sieciach, w których niewielka ilość węzłów łączy się z dużą ilością pozostałych wierzchołków znacznie przewyższając liczbą wychodzących gałęzi znakomitą większość węzłów? Żadna z przedstawionych do tej pory koncepcji budowy sieci nie przewiduje takiego przypadku. Wraz z rozwojem sieci internetowej model takiego zachowania został szybko zauważony i poddany analizie. Popularne wyszukiwarki takie jak google.pl w sieci Web zostały zinterpretowane jako tzw. centra kontrastując z rzadko odwiedzanymi przez internautów stronami prywatnymi. Sieci tego typu ze względu na bardzo zróżnicowaną i złożoną budowę zyskały miano bezskalowych.

Szerzej wymienionym strukturom sieci przyjrzymy się w kolejnych rozdziałach, ponieważ poznanie ich atrybutów pozwoli nam na dostrzeżenie specyfiki innych, kompleksowych układów sieciowych.

## 1.2 Cel

Przybliżenie struktury teorii grafów od strony metodologii budowy sieci, na przykładzie danych rzeczywistych to jeden z celów tej pracy. Zdecydowałam się na realizację tego zamierzenia ze względu na szerokie wykorzystywanie wiedzy o budowie sieci w wielu dziedzinach życia na Ziemi.

Dodatkowo w mojej pracy skupię się na badaniu sieci pod względem jej modułowości. Dzięki wykonanym przeze mnie programom można będzie określić przynależność do konkretnych grup (w tym przypadku np. przynależność do klas użytkowników portalu [nasza-klasa.pl](http://nasza-klasa.pl)) jedynie na podstawie analizy wykresów wygenerowanych przez program. Inspiracją do budowy programu i powodem zainteresowania tym tematem był artykuł „Diffusion on complex networks: a way to probe their large-scale topological structures” autorstwa Simonsena i in. [3]. W ten sposób bez wizualizacji sieci za pomocą węzłów i połączeń między nimi możemy zaobserwować przykładowo przynależność użytkownika do grupy dzięki specjalnie do tego celu stworzonym diagramie. Większość analiz dotyczących sieci będzie możliwa tylko i wyłącznie dzięki pomocy komputera. Z tego względu sieci będą przedstawiane w zrozumiałym dla niego języku, czyli w układzie macierzowym, przedstawiającym połączenia w sieci za pomocą zer i jedynek. Szczegółowo macierzowy wymiar sieci opisze w następnych rozdziałach.

Będę się również starała udowodnić, iż wygenerowanie sieci „podobnych” (sieci posiadających taką samą ilość węzłów i analogiczną liczbę stopni wierzchołków wychodzących z węzłów) nie łączy się z przejściem przez nie własności sieci wyjściowej. Sieci podobne różnicuje pod względem przebiegu algorytmu łączenia się z kolejnymi wierzchołkami. Dzięki temu będziemy mogli zaobserwować wpływ zmiany własności sieci podczas modyfikacji algorytmu postępowania przy jej budowie.

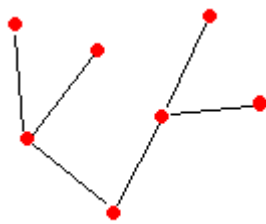
Aby przybliżyć i zbadać strukturę sieci w pracy zbadam sieci tworzone na bazie portalu [nasza-klasa.pl](http://nasza-klasa.pl). Na podstawie wygenerowanych, różniących się między sobą sieci będę starała się zaobserwować modułowość (przynależność do klas, szkół) obserwując wyłącznie wykresy generowane przez program „Analiza sieci”. Zaobserwowanie modułów w sieci jest dość trudne, ze względu na duże problemy w zobrazowaniu sieci w postaci węzłów i krawędzi je łączących. Zwykle po próbie narysowania takiej sieci wykres na ekranie komputera jest nieczytelny, a zatem i bezużyteczny do jakichkolwiek analiz. Jednocześnie algorytmy komputerowe, rysujące sieć w sposób jak najbardziej przejrzysty, czyli z jak najmniejszą ilością przecinających się węzłów mają zbyt dużą złożoność obliczeniową, aby można było je efektywnie wykorzystać do analiz sieci o dużej liczbie wierzchołków i krawędzi. Spróbuję również odpowiedzieć na pytanie jak można przewidzieć przynależność do grup, klas, kółek zainteresowań jedynie na podstawie diagramów opartych o struktury sieci znajomości.

### **1.3 Zakres pracy**

Zrozumienie sensu tematów poruszanych w pracy wymaga odpowiedzi na pytanie: dlaczego identyfikacja struktur sieci jest taka ważna?

W mojej pracy skupię się szczególnie na identyfikacji struktur sieci złożonych, w których często mamy do czynienia z modułowością i skomplikowaną strukturą, co z kolei wymaga obszerniejszej analizy niż ograniczenie do opierania się o wiedzę zgromadzoną na temat sieci prostych pod względem budowy (grafy proste, sieci regularne). W tym celu krokowo będę starała się przybliżyć strukturę sieci kładąc nacisk na rozpatrywanie ich pod względem występowania tzw „grup znajomych” w sieciach.





*Rys. 1.3.1 Przykład grafu prostego.*

Wszyscy jesteśmy powiązani i tworzymy układy różnych relacji towarzyskich, biznesowych, politycznych. Nasi znajomi mają swoich znajomych stanowiących często jedynie część w innych sieciach zależności. Jakościowo naturę tych sieci dosyć dobrze rozpoznano już kilkadziesiąt lat temu. Na początku lat siedemdziesiątych badacze z Rand Corporation, na podstawie analizy ruchów kontrkulturowych i organizacji terrorystycznych stworzyli model SPIN (Segmented Polycentric Ideological Network – wielosegmentowa wielocentryczna sieć scalona wspólną ideologią). Wyjaśnia on, dlaczego tak trudno skutecznie walczyć z terrorystami i podziemiem gospodarczym – są to organizacje pozbawione hierarchicznej struktury, proces decyzyjny rozproszony jest w sieci relacji członków grupy a eliminacja jednego węzła nie niszczy całości.[1]

Jak wynika z powyższego tekstu znaczenie poznania struktur w ujęciu sieciowym jest ogromne a korzyści płynące z prawidłowej analizy nieocenione. Jednym z praktycznych zastosowań znajomości topologii sieci może być również istotne przyspieszenie przepływu informacji, bez konieczności znacznej przebudowy sieci. Zainteresowani mogą być też epidemiolodzy - model „Małego Świata” znakomicie tłumaczy rozprzestrzenianie się chorób zakaźnych [2].

Posiadanie informacji o topologii sieci ma szerokie zastosowanie w praktyce. Przedstawianie rzeczywistych kontaktów i znajomość ilości połączeń odchodzących od każdego węzła generowanej na tej podstawie sieci jest w dzisiejszych czasach coraz bardziej ceniona chociażby ze względu na coraz częstsze wykorzystanie zasobów zgromadzonej wiedzy podczas rozwiązywania problemów w ujęciu strategicznym i planistycznym. Wiedza o tzw. modułach (klastrach) jest bardzo przydatna chociażby w tworzeniu skutecznych strategii marketingowych. Moduły możemy obrazować jako skupiska ludzi, którzy są w bardzo zacieśnionych stosunkach. Informacja, która trafi do takich osób tworzących zacieśniony moduł może w jego obrębie bardzo długo przepływać i nie wydostawać się na zewnątrz (inne części sieci). Jeśli natomiast wiemy, którzy ludzie są tzw. centrami, to jest to często równoznaczne z informacją jak szybko dotrzeć do największej liczby osób, a co za tym

idzie wiemy jak i kogo najpierw zainteresować promocją produktu aby najefektywniej go wypromować.

W związku z tym zakres pracy będzie obejmował:

- 1) Wprowadzanie w zagadnienia teoretyczne dotyczące topologii sieci.
- 2) Wyjaśnienie użyteczności badań nad procesami dyfuzji w sieciach.
- 3) Stworzenie programu do pobierania danych z portalu internetowego.
- 4) Opis i analiza wykorzystanych algorytmów.
- 5) Utworzenie aplikacji do rysowania, analizowania sieci i identyfikacji struktur pod względem ich złożoności.
- 6) Przegląd i analiza budowy wygenerowanych sieci, identyfikacja struktur
- 7) Identyfikacja modułów, centrów w sieciach złożonych (przewidywanie przynależności do grup „dobrych znajomych”)

## **2. Podstawowe pojęcia i definicje**

### **2.1 Podstawowe pojęcia dotyczące sieci**

Dla ułatwienia zrozumienia pracy przedstawię poniżej parę definicji pojęć, które często wykorzystywane są w pracy do opisu zjawisk, reguł i analiz. Definicje opracowałam na podstawie prac [3], [4], [5], [6] i [7].

**Siec** – każdy system złożony z elementów połączonych relacjami.

**Graf** – matematyczne ujęcie sieci.

**Stopień wierzchołka** – liczba wychodzących z wierzchołka krawędzi.

**Teoria grafów** – dziedzina matematyki zajmująca się sieciami.

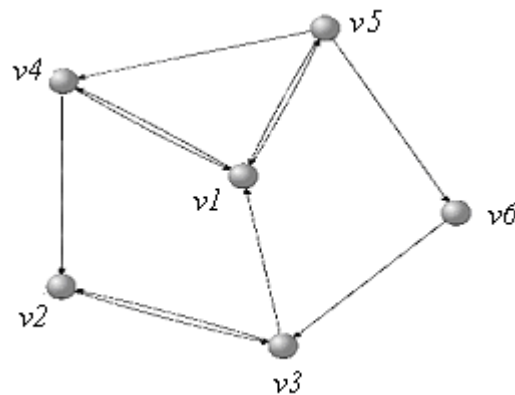
**SNA (Social Network Analysis)** – analiza sieci społecznych. Dyscyplina, która czerpie z teorii grafów na gruncie nauk społecznych. Jej przedmiotem są sieci tworzone przez różnego rodzaju „obiekty” społeczne – są nimi najczęściej ludzie, ale także grupy, firmy czy nawet państwa. Stanowią one węzły sieci i są ze sobą połączone relacjami, także o charakterze społecznym. Relacje mogą być symetryczne (np. komunikowanie się, współpraca - z definicji zawsze odwzajemnione) albo skierowane (np. pytanie o radę – które są jednostronne).

**Centrum** – wierzchołek mający stosunkowo znacznie więcej połączeń względem innych węzłów. Usunięcie centrum z grafu wiąże się z dużym rozspójnieniem grafu.

**Wierzchołki incydentne** – wierzchołki sąsiednie, czyli takie które łączą krawędź.

**Macierz incydencji** – macierz wymiaru  $n \times m$ , której wyraz o indeksach  $i, j$  jest równy 1, jeśli wierzchołek  $i$  jest incydentny z krawędzią  $j$ , i równy 0 w przeciwnym wypadku. Zapis macierzowy sieci jest jednym z najczęstszych sposobów przekazywania informacji o sieci do komputera.

	v1	v2	v3	v4	v5	v6
v1		0	0	1	1	0
v2	0		1	0	0	0
v3	1	1		0	0	0
v4	1	1	1		0	0
v5	1	0	0	1		1
v6	0	0	1	0	0	



Rys 2.1.1 Przykład opisu sieci za pomocą macierzy.

**Wektor własny** – wektor własny macierzy  $A$ , to każdy taki wektor  $\vec{x}$ , dla którego zachodzi równość:

$$A \cdot \vec{x} = \lambda \cdot \vec{x}$$

gdzie  $\lambda$  to wartość własna macierzy.

**Klika (moduł, klaster)** – rodzaj spójnej grupy w grafie, grupa co najmniej 3 węzłów, z których każdy ma połączenie z każdym innym.

**Klasteryzacja** – sposób grupowania (ang. clustering), interpretowane jako rozmieszczenie blisko siebie obiektów w jakiś sposób ze sobą powiązanych, a powstała struktura fizyczna określana jest jako klastery lub grono, np. celem grupowania obiektów na dysku jest zwiększenie efektywności przetwarzania zapytań przez zmniejszenie ilości operacji WE/WY koniecznych do odczytania potrzebnych danych.

**Współczynnik gronowania (clustering coefficient)** – służy do szacowania, ile sąsiadów danego wierzchołka jest połączonych każdy-z-każdym. Jest to iloraz liczby krawędzi pomiędzy sąsiadami danego wierzchołka oraz liczby krawędzi, jaki miałby graf pełny składający się z tych sąsiadów. Mówi on, jak bardzo „znają się” sąsiedzi danego wierzchołka. Wysoki współczynnik gronowania świadczy o tym, że moduł, do którego należy wierzchołek jest „gęsty” (większość wierzchołków w module jest ze sobą połączona „każdy z każdym”).

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)} : v_i, v_k \in N_i, e_{ij} \in E$$

$C_i$  – współczynnik gronowania

$\{e_{jk}\}$  – zbiór krawędzi pomiędzy sąsiadami wierzchołka  $i$

$k_i$  – stopień wierzchołka  $i$

**Średni współczynnik gronowania** – średni współczynnik klasteryzacji po wszystkich wierzchołkach grafu. Przyjmuje wartości pomiędzy 0 i 1.

**Średnia odległość między węzłami** – Średnia długość ścieżki w sieci mówi przez ile wierzchołków należy średnio przejść, aby dojść od jednego wierzchołka do dowolnego innego. Będziemy ją oznaczać literą  $l$ .

## 2.2 Sieci i ich własności (sieci złożone, struktury sieci)

Przejsie do analizy topologii sieci wymaga zapoznania się z kilkoma podstawowymi rodzajami sieci. Z wieloma z nich mamy do czynienia na codzień, ich struktury przewijają się w naszym życiu choć nie zawsze zdajemy sobie z tego sprawę. Przyjrzyjmy się zatem kilku szczególnym rodzajom sieci, którym podporządkowana jest znaczna ilość sieciowych struktur w naszym świecie.

### 2.2.1 Sieć regularna

Sieć regularna jest najprostszym modelem, który jednak ma zastosowanie w niewielu złożonych układach. Cechą charakterystyczną takiej sieci jest jednostajny rozkład połączeń. W rzeczywistości z takimi sieciami mamy do czynienia dość rzadko, ponieważ ogólnie idealne i równomierne struktury są rzadziej spotykane w naturze.

Sieć regularna jest podobna do krystalicznej. Atomy w kryształach łączą się wiązaniami właśnie ze swymi najbliższymi sąsiadami. Fachowo mówi się o takiej sieci, że ma wysoki współczynnik gronowania. Niestety, sieć regularna ma jeden poważny defekt – długą średnią drogę między węzłami. Sieć regularna charakteryzuje się rozkładem jednostajnym stopnia węzłów [9].

Sieć regularna cechują się:

- wysokim współczynnikiem gronowania,
- długą średnią odległością pomiędzy węzłami,

Sieć regularna nie nadaje się do modelowania bardziej złożonych koneksji w społeczeństwie, a także wielu kompleksowych struktur występujących w naturze.



Rys 2.2.1 Schemat jednowymiarowej sieci regularnej. Węzły łączą się z sąsiadami pierwszego i drugiego rzędu. Źródło: [8].

### 2.2.2 Sieć losowa

Model sieci losowej zwany też modelem Erdösa–Rényi (ER) to sieć  $N$  węzłów, gdzie każda para węzłów jest połączona krawędzią z prawdopodobieństwem  $p$ , co daje graf o około  $pN(N-1)/2$  losowo rozmieszczonych krawędziach. Rozkład ilości połączeń tej sieci,  $P(k)$ , gdzie  $k$  jest stopniem wierzchołka, jest rozkładem Poissona, co oznacza, że większość węzłów ma w przybliżeniu tę samą liczbę odchodzących od niego krawędzi. Wierzchołki, które mają inny stopień niż  $k$  są bardzo rzadkie - im większe odchylenie od  $k$ , tym mniej takich wierzchołków. Współczynnik gronowania danego węzła jest niezależny od stopnia węzła. Stąd funkcja  $C(k)$  (czyli zależność współczynnika gronowania od stopnia węzła) jest prostą poziomą, co oznacza, że posiada własności sieci Małego Świata [9].

Podsumowując, w sieci losowej węzły z liczbą połączeń znacząco różną od średniej trafiają się rzadko. Prawdopodobieństwo przyłączenia kolejnego węzła w sieci losowej maleje wykładniczo dla wierzchołków o coraz większym stopniu. Ze względu na tę własność centra (węzły o znacznie większej niż przeciętna liczbie połączeń) nie mają prawa istnieć w takiej sieci.

Cechami charakterystycznymi sieci losowych jest:

- krótka średnia droga między węzłami
- niski współczynnik gronowania.

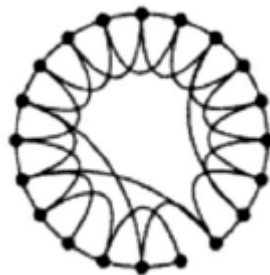


Rys.2.2.2 Schemat budowy sieci losowej. Źródło: [8].

### 2.2.3 Sieci Małego Świata

W celu opisu sieci społecznych dwaj matematycy, Watts i Strogatz, przedstawili nowy model sieci [10]. Nowa sieć miała łączyć w sobie pozytywne cechy sieci regularnej i losowej: wysoki współczynnik gronowania (sieć regularna) i krótką średnią drogę połączeń między węzłami (sieć losowa).

Budowa sieci Małego Świata okazuje się bardzo prosta. Jeżeli tylko trochę zaburzymy sieć regularną, tzn. jeśli np. 1% połączeń przerwiemy i zastąpimy je połączeniami losowymi, to otrzymana struktura będzie się świetnie nadawać do opisu np. zjawisk socjologicznych [999].



Rys.2.2.3 Schemat budowy jednowymiarowej sieci Małego Świata. Źródło: [8].

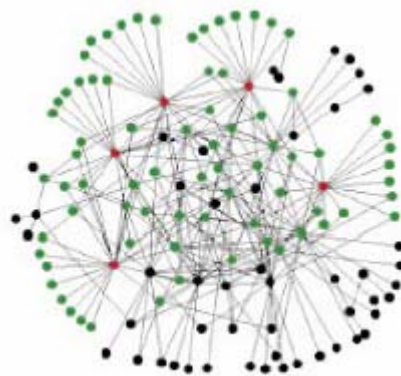
Sieć Małego Świata spełnia dwa warunki, które kwalifikują ją jako narzędzie do modelowania zachowań społecznych:

- średnia odległość między węzłami jest krótka (szybki przepływ informacji w sieci)
- wysoki współczynnik gronowania (istnienie zwartych grup połączeń w sieci)

Model Małego Świata nie może jednak posłużyć do obrazowania sieci z tzw. „centrami”, czyli nie bierze pod uwagę osób, które zawarły znacznie więcej znajomości od swoich kolegów (rozkład wiązań nie jest potęgowy). Taki warunek spełnia za to inny rodzaj sieci, który przedstawię w kolejnym punkcie.

### 2.2.4 Sieci bezskalowe

Skupienie uwagi naukowców na węzłach o zdecydowanie przewyższającej i znacznie odbiegającej pod względem ilości połączeń od pozostałych grup węzłów w 1999 roku zaowocowało zdefiniowaniem nowego pojęcia w teorii o sieciach. Główną przyczyną powstania nowego algorytmu opisującego grupę sieci było coraz szersze zainteresowanie wokół opracowania algorytmu rozrastania się sieci internetowej. Dwaj fizycy, Barabási i Albert, zaproponowali nowy przepis na tworzenie sieci, nazwanych później bezskalowymi [3]. Zauważyli oni, iż sieci, określane przez nich jako bezskalowe, mimo swojej złożonej struktury rozrastają się w sposób nieprzypadkowy. Kolejne węzły przyłączają się z większym prawdopodobieństwem do tych, które posiadają większą ilość połączeń. Stąd też szybko tworzą się tzw. centra czyli węzły o ilości wychodzących wiązań znacznie przewyższającej stopnie większości pozostałych [9].



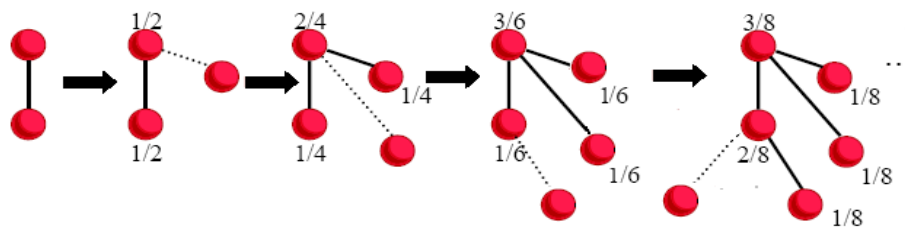
Rys.2.2.4 Przykład sieci beskalowej. Źródło:[11].

Barabási nie poprzestał jednak na Internecie i zaczął się zastanawiać, czy odkryty przez niego model matematyczny nadaje się do analizy innych sieci? Na przykład do opisu rozkładu bogactwa. Podobnie jak ruch w Internecie, również majątek koncentruje się w kilku węzłach – w rękach superbogatyh ludzi. Fenomen ten opisywała w przybliżeniu licząca sto lat zasada Pareto, mówiąca, że 80 proc. bogactwa jest w rękach 20 proc. ludzi. Prace Barabásiego nadały tej zasadzie rygor matematyczny, a dalsze badania odkryły kolejne prawidłowości [1].

Sieć bezskalowa może rosnać w nieskończoność, współczynnik gronowania ma średnio blisko pięć razy większy niż sieć losowa (czyli dobrze oddaje fakt trzymania się



znajomych w paczce), a średnia odległość między węzłami jest dość krótka (świat Barabásiego jest więc ‘mały’). Taką architekturą rozmieszczenia w przestrzeni charakteryzują się społeczeństwa ale posiada ją również wiele innych układów: pajęczyna stron WWW połączona linkami, fizyczna struktura Internetu, routery połączone światłowodami, sieć poczty elektronicznej, sieć współpracy między naukowcami i powiązań gospodarczych. Identyczną budowę ma nawet sieć zależności między białkami regulującymi działanie komórki [9].



*Rys.2.2.4 Proces tworzenia sieci bezskalowej (w sieci bezskalowej kolejne węzły z większym prawdopodobieństwem przyłączają się do tych węzłów, które aktualnie mają najwięcej połączeń). Źródło:[9].*

Sieci bezskalowe charakteryzują się rozkładem potęgowym. Prawdopodobieństwo, że węzeł ma stopień  $k$  jest proporcjonalne do  $k^{-\gamma}$  gdzie  $\gamma$  to wykładnik stopnia wierzchołka. Prawdopodobieństwo, że węzeł ma wysoki stopień jest statystycznie bardziej znaczące (większe) niż w grafie losowym. Często sieci takie mają względnie małą liczbę węzłów o wysokim stopniu (centra, huby). W modelu Barabásiego–Alberta prawdopodobieństwo, że nowy węzeł o stopniu  $M$  dodany do sieci podłączy się do istniejącego już węzła  $I$  wynosi  $k_I / \sum k_J$ , gdzie  $k_I$  to stopień węzła  $I$  a  $J$  oznacza kolejne indeksy węzłów sieci. Taka sieć ma rozkład potęgowy o wykładniku  $\gamma = 3$ . Rozkłady ilości połączeń na wykresie logarytmicznym są liniami prostymi. W tym modelu sieci nie charakteryzuje modułowość, więc współczynnik gronowania  $C(k)$  jest niezależny od  $k$ . Sieci bezskalowe z wykładnikiem stopnia  $2 < \gamma < 3$  (zakres obserwowany w wielu biologicznych i niebiologicznych sieciach) są dodatkowo „ultra-małe”. Średnia odległość między węzłami w takich sieciach to  $l \sim \log \log N$ , co jest znacząco mniejsze niż złożoności klasy  $\log N$ , która charakteryzuje losowe Małe Światy [9].

Własności sieci bezskalowych:

- węzły o dużym stopniu wierzchołka występują rzadko.
- odporne na przypadkowe uszkodzenia

- mało odporne na zaplanowane ataki na centra. Skoordinowany atak powoduje natychmiastowe rozspójnienie sieci (dekompozycja sieci na mniejsze fragmenty przez usunięcie powiązań pomiędzy ścisłymi modułami)

Ze względu na częste pokrywanie się struktury sieci bezskalowej ze strukturą **sieci społecznej** zdecydowałam się krótko opisać również tę ważną grupę sieci. W ostatnich latach zwraca się na nią szczególna uwaga, ponieważ znajomość topologii sieci społecznych znacznie wspomaga przewidywanie rozprzestrzeniania się informacji wśród grup społecznych. Pozwala to przewidzieć zachowanie konsumentów, budować efektywne strategie marketingowe, unikać sytuacji krytycznych (rozpad kluczowych połączeń).

Prawidłowy model sieci społecznej powinien posiadać trzy następujące cechy:

- wysoki współczynnik gronowania (osoby mieszkające na jednym osiedlu zwykle znają się wzajemnie),
- niewielką średnią odległość między węzłami (informacje w sieci społecznej rozchodzą się szybko),
- potęgowy rozkład liczby połączeń (warunkuje istnienie centrów, czyli osób które posiadają znacznie więcej znajomych niż inni).

Budowa sieci społecznych najczęściej pokrywa się z topologią sieci bezskalowych, mają również wiele wspólnego z budową sieci Małego Świata.

Żeby podkreślić rosnącą potrzebę pogłębiania wiedzy o sieciach warto wspomnieć o zatrudnieniu przez zespół Yahoo Duncana Wattsa, jednego z czołowych badaczy sieci szczególnie od strony socjologicznej. Badania nad modelem Małego Świata prowadzone przez Wattsa w 1998 roku rozwinęły zainteresowanie bardziej złożonymi sieciami i sprawiły, że fizycy z większym zacięciem zaczęli interesować się badaniami społecznymi i analizą sieci. Sam Watts, który robił doktorat z fizyki, był do tej pory profesorem socjologii. Taki transfer z uniwersytetu do biznesu jest ostatnio coraz częściej spotykany [12].

## **2.3 Wykorzystanie badań nad topologią sieci**

Każdą sieć można zinterpretować jako graf, w którym wierzchołki połączone są krawędzią, jeśli spełniają daną relację. Relacją może być np. relacja znajomości, a wierzchołki można interpretować jako osoby. Powstała w ten sposób sieć jest siecią znajomości. Jeśli uda nam się zidentyfikować moduły w takiej sieci oraz tworzące je wierzchołki, możemy określić, które osoby tworzą grupy „starych dobrych znajomych”. Podobnie, jeśli wierzchołki odpowiadają autorom, a krawędzie łączą tych autorów, którzy często razem publikują prace, moduły w takiej sieci można interpretować jako grupy specjalistów z tej samej dziedziny. Dzięki temu, w oparciu o analizę dyfuzji w sieciach, bez początkowej wiedzy na temat topologii sieci, możemy uzyskać wiele informacji na temat relacji pomiędzy osobnikami.

## **2.4 Zastosowanie wiedzy o centrach w sieciach**

Ponieważ przytłaczająca liczba węzłów łączy się z centrami, nawet mało zaraźliwy wirus wpuszczony w taką strukturę sieciową natychmiast zostanie przekazany do olbrzymiej ilości innych węzłów. Zaatakuj jedno z centrów, a następnie kolejne i opanuj cały system. Choć taka topologia może być dość niebezpieczna, odpowiednio wykorzystana może także być sojusznikiem. Wiedza o centrach pomoże skutecznie i szybko zwalczyć epidemie poprzez odpowiednią dystrybucję szczepionek (uodpornianie osób posiadających dużą ilość kontaktów w pierwszej kolejności). Oczywiście wiąże się z tym także aspekt etyczny, zwłaszcza gdyby szczepionek miało nie wystarczyć dla wszystkich.

Specjaliści od marketingu z kolei zyskaliby szybsze zainteresowanie nowymi towarami kierując kampanie reklamowe do osób będącymi centrami w sieciach społecznych, sieci e-mail oraz sieci rozmów telefonicznych itd. [14]. Osoby zajmujące się marketingiem dobrze wiedzą, że skierowanie swojej uwagi na ludzi, którzy posiadają wielu znajomych i jednocześnie nie zamykają przed zawieraniem nowych znajomości, jest często gwarancją sukcesu efektywnego skutku reklamy.

## 2.5 Wykorzystanie badań nad przebiegiem procesów dyfuzji

### 2.5.1 Co nazywamy procesem dyfuzji?

Kolejne z zagadnień dynamiki w sieci to problem rozprzestrzeniania się (dyfuzji) w sieciach różnych czynników – chorób, innowacji, plotek, mód, zachowań. Podobnie jak w przestrzennej dynamice wpływu społecznego, także i tu mamy do czynienia z progowym mechanizmem rządzącym zmianami zachodzącymi w jednostce. Nie tylko w przypadku chorób (oraz w pewnych przypadkach marketingu wirusowego) jest tak, że to, czy się zarazimy, nie jest zależne wyłącznie od procentowego udziału kontaktów w liczbie zarażonych – istotą „zaraźliwości” (*social contagion*) jest addytywny wpływ otoczenia na nasze zachowanie [5].

### 2.5.2 Zastosowanie wiedzy o procesach dyfuzji i ich rozpoznanie w układzie

Określenie topologii sieci wiąże się z identyfikacją struktur takich jak moduły (klastry). Aby znaleźć wierzchołki tworzące moduły, można posłużyć się analizą procesu dyfuzji w sieci, polegającego na symulacji błądzenia losowego po wierzchołkach sieci. Jeśli umieścimy na wierzchołkach pewną liczbę pionków („random walkers”) przeskakujących losowo z wierzchołka na wierzchołek, spodziewamy się, że jeśli wejdą one na wierzchołki należące do modułu, będą odwiedzać je z dużą częstotliwością. Taki proces dyfuzji, gdzie prawdopodobieństwo przejścia z jednego wierzchołka na dowolny inny z nim połączony jest jednorodne oraz niezmiennie w czasie, jest procesem Markowa. Częstotliwość odwiedzania danych wierzchołków w takim procesie można oszacować obliczając wartości i wektory własne macierzy przejścia tego procesu. Wartości własne pozwalają oszacować jak szybko maleje częstotliwość odwiedzania danego wierzchołka. Wektor własny związany z danym wierzchołkiem opisuje jak często odwiedzane są jego krawędzie. Interesują nas wierzchołki, których częstotliwość odwiedzania zanika w zbliżonym, ale bardzo małym tempie. Można się spodziewać, że w większości przypadków takie wierzchołki muszą być odwiedzane przez te same „pionki” błądzące (oscylujące) wewnątrz tego samego modułu. A więc aby wyznaczyć wierzchołki tworzące moduły, należy obliczyć wektory i wartości własne ( $C_i$ ) macierzy przejścia dla dwóch jak największych (mniejszych od 1) wartości własnych (*alfa*) i porównać je [3].

*W celu porównania tworzymy wykres, na którym zaznaczamy kolejne wierzchołki (i) interpretując ich wektory własne jako współrzędne  $x$  i  $y$  dla pierwszej i drugiej wartości własnej. Te wierzchołki, których częstotliwość odwiedzania zmienia się w tym samym tempie (a więc zapewne należą do tego samego modułu) będą układały się na liniach prostych na tym wykresie. Stąd wierzchołki leżące na linii prostej można zakwalifikować jako należące do jednej grupy powiązanych ze sobą osób [9].*

### **2.5.3 Znaczenie badań nad procesami dyfuzji**

Dzięki szczegółowemu poznaniu specyfiki procesów dyfuzji zachodzących w sieci można przewidzieć topologię sieci po jej dekompozycji, czyli jeśli dojdzie do rozpadu sieci na kilka mniejszych po usunięciu tzw. „słabych więzi” między modułami. Można z dużym prawdopodobieństwem stwierdzić, które węzły po usunięciu owych połączeń będą przynależą do nowopowstałych sieci.

Posiadanie wiedzy o procesach dyfuzji w sieci jest jednoznaczne z posiadaniem informacji o występujących w niej modułach. Poszukiwanie klastrów (grup współników) jest trudne ze względu na problemy w przedstawieniu większych sieci w przejrzystej formie graficznej, w której można by zaobserwować skupiska połączeń. Współczynnik klasteryzacji, mówiący o gęstości krawędzi, ułatwia poznanie topologii sieci pod względem występujących w niej zagnieżdżeń ale jednocześnie pozwala jedynie na bardzo ogólny opis topologii pod tym względem. Przykładowo mając dane z portalu [nasza-klasa.pl](http://nasza-klasa.pl) (portal, w którym użytkownik może gromadzić zdjęcia i informacje o znajomych, którzy z kolei posiadają innych znajomych z klasy i nie tylko) wiedzę o sieci znajomości i procesach dyfuzji w niej zachodzących bez informacji o przynależności poszczególnych osób do klas, możemy odbudować tę informację (o składzie klas). Możemy też określić, którzy autorzy specjalizują się w tych samych dziedzinach pomimo, iż nie znamy tematyki dzieł a jedynie posiadamy wiedzę na temat tego kto z kim najczęściej współtworzy publikacje naukowe.

Wiedza na temat dyfuzji o rozprzestrzeniających się opiniach, wirusach, sprzedających się produktach i rozwijających się technologiach ma szczególne znaczenie w dzisiejszym świecie, w którym odkrycie newralgicznych punktów może zdecydować o sukcesie firmy, pomoże ulokować reklamę w najodpowiedniejszym miejscu lub powstrzymać rozprzestrzeniającą się epidemię (mając sieć znajomych możemy zbadać przyrost

„zarażonych” węzłów w czasie) a także przewidzieć układ sił po rozpadzie ugrupowania politycznego.

### **3. Technologie użyte do zbudowania programu „Analiza sieci”.**

Zbadanie właściwości i struktury różnego rodzaju sieci złożonych ułatwił mi zaprojektowany przeze mnie program. Przy jego budowie korzystałam z narzędzi takich jak makra w Excelu do konwersji danych aby umożliwić pracę na plikach programowi grabber.pl, który z kolei służy do pobierania danych z portali (w tym przypadku nasza-klasa.pl). Efektem działania tych dwóch programów są dane przygotowane w odpowiedniej formie do pobrania przez program „Analiza sieci” wykonany za pomocą narzędzia programistycznego Matlab. Dla osób chcących zapoznać się z działaniem aplikacji „Analiza sieci” proponuje przeczytanie informacji zamieszczonych w sekcji Dodatek, w której omawiam podstawowe pojęcia programistyczne, które warto poznać przed obejrzeniem kodów programów.

#### **3.1 Matlab jako podstawowe narzędzie programistyczne<sup>\*</sup>.**

Dlaczego do zbudowania programu „Analiza sieci” wykorzystałam pakiet Matlaba? Matlab jest środowiskiem obliczeniowym przeznaczonym dla inżynierów i naukowców, umożliwiającym przeprowadzanie obliczeń matematycznych, analizy numerycznej, wizualizacji otrzymanych wyników (2D, 3D), jak również tworzenie algorytmów i programów. Język jest wygodny w użyciu, co sprawia, że opracowanie algorytmów jest prostsze niż w przypadku takich języków programowania jak C++ dlatego wybrałam go jako główne narzędzie programistyczne. Matlab odpowiadał mi również pod względem funkcjonalności:

- obliczenia numeryczne do szybkiego generowania wyników
- grafika do wizualizacji i analizy danych
- interaktywny język i środowisko programistyczne
- narzędzia do budowy własnego GUI

---

<sup>\*</sup>Na podstawie [15]

integracja z zewnętrznymi aplikacjami składającymi się z komponentów, np. Excel.

import danych z plików i urządzeń zewnętrznych (dodatkowo dostęp do baz danych i kolejnych urządzeń)

Matlab pozwala na wykonanie w sposób przejrzysty prostych operacji na macierzach. Wadą oprogramowania jest jednak powolność podczas wykonywania obliczeń na macierzach wymagających dużej mocy obliczeniowej.

## 4. Program „Analiza sieci”

### 4.1 Formaty danych do programu „Analiza sieci”

Do analizy wykorzystane zostaną między innymi sieci znajomości, pochodzące z portalu „nasza-klasa.pl”. Dane pobierane są z portalu i przygotowywane za pomocą zestawu dwóch programów napisanych w Perlu :

1) „*grebber.pl*” – działanie programu polega na zalogowaniu do portalu „nasza-klasa.pl” jako fikcyjny użytkownik, przetwarzaniu danych zwracanych przez portal strony HTML oraz wyciąganiu informacji na temat zarejestrowanych w nim osób tworząc:

- listę znajomych
- listę szkół (klas) do której należy dana osoba

Program, po zalogowaniu, pobiera z pliku „start.txt” listę osób (oraz zakres skanowania ich znajomych), które będzie przetwarzać. Następnie rekursywnie analizuje profil kolejnych osób, ściągając informacje o jej znajomych i szkołach. Informacje te zapisywane są w plikach „znajomi.przed.txt” oraz „klasy.przed.txt”.

Każda linijka pliku „znajomi.przed.txt” składa się z 2 liczb oddzielonych przecinkiem: A,B. Wiersze te są numerami profili osób zarejestrowanych w portalu, które się znajdują. Natomiast wiersze w pliku „klasy.przed.txt” składają się z 2 liczb oddzielonych przecinkiem: A,B. W pliku tym A jest numerem profilu osoby a B jest identyfikatorem szkoły (klasy) do

---

której dana osoba należy. Plik „start.txt” składa się z linijek, w których każda jest postaci :  $A,B,C$ , gdzie  $A$  jest numerem profilu osoby, z której znajomych będziemy budować wierzchołki i krawędzie grafu.  $C$  oznacza maksymalny stopień wierzchołka (maksymalną liczbę bezpośrednich znajomych danej osoby, jaką program pobierze) natomiast  $B$  jest „głębokością” rekursji.

2) „renumerator.pl” – konwertuje profile i identyfikatory z plików „znajomi.przed.txt” oraz „klasy.przed.txt” zastępując je kolejnymi liczbami. Ponieważ numery profili i identyfikatory szkół (klas) są dużymi liczbami, nie można ich użyć bezpośrednio w narzędziu służącym do analizy sieci. Program numerując kolejne osoby i szkoły (klasy) i zapisując wyniki w plikach „znajomi.po.txt” oraz „klasy.po.txt” umożliwia wczytanie tych danych bezpośrednio przez program analizujący sieci.

## 4.2 Opis korzystania z programu

W oknie programu możemy wyróżnić 3 sektory:

1) Ustawienia – zawiera pola, w których należy wpisać ścieżki do plików, zawierających dane o sieciach i ewentualnie klasach, oraz pozwala wybrać wartości własne, na których będzie opierała się analiza. Pozwala też wybrać klasę, której wierzchołki będą zaznaczone na wykresie kolorem czerwonym.

2) Akcje i statystyki – Wyświetla informacje o sieci: ile zawiera wierzchołków, jaki ma współczynnik gronowania oraz ewentualnie ile wierzchołków należy do tej samej wybranej klasy.

3) Wykres – w zależności od wybranej akcji (rysuj, rozkład) i ewentualnie klasy, rysowany jest wykres obrazujący przebieg modułowości sieci bądź rozkład sieci..

Aby zanalizować sieć, należy najpierw w polu „plik znajomych” podać nazwę pliku, w którym przechowywana jest informacja na temat krawędzi grafu. Każda linia tego pliku składa się z dwóch liczb oddzielonych przecinkiem postaci:  $A,B$  i odpowiada krawędzi łączącej wierzchołek  $A$  oraz  $B$ .

Jeśli znana jest nam informacja na temat przynależności wierzchołków do klas (modułów, podmodułów, szkół) możemy podać adres pliku z tą informacją w polu „plik klas”. Każda linia tego pliku musi być postaci:  $A,B$ , gdzie  $A$  oznacza wierzchołek, a  $B$  klasę,



do której należy wierzchołek A. Następnie wybieramy akcję „Wczytaj”. Po wczytaniu pliku, pojawia się informacja o liczbie wierzchołków w sieci oraz współczynniku gronowania całej sieci.

Wybierając akcję „Rysuj rozkład” otrzymamy wykres przedstawiający rozkład sieci. Oś pozioma oznacza stopień wierzchołka, oś pionowa liczbę wierzchołków o danym stopniu. Wybierając akcję „Rysuj” tworzony jest wykres, na którym umieszczone są wierzchołki. Te wierzchołki, które układają się w linie proste, stanowią klastry w sieci. Jeśli dostępna jest informacja o klasach, wierzchołki wchodzące w skład wybranej klasy zostaną dodatkowo zaznaczone na czerwono.

Inicjując akcję za pomocą przycisku „Sieć losowa” wygenerowana zostanie losowa sieć o takim samym rozkładzie, co sieć wyjściowa. Analogicznie wywołanie akcji „Sieć zachłanna” wygeneruje deterministycznym algorytmem sieć o takim samym rozkładzie.

### **4.3 Algorytmy stosowane w programie**

W programie „Analiza sieci” wykorzystałam dwa algorytmy do budowania sieci na podstawie rozkładu sieci stworzonej przez program „grebber.pl” z portalu nasza-klasa.pl lub innej dowolnie dobranej. Pierwszy z nich opiera się na losowym wyborze kolejnych węzłów, w drugim algorytmie sugeruję się stopniem wierzchołków i tworzenie połączeń rozpoczynam od węzłów z największym stopniem wierzchołka.

#### ***4.3.1 Algorytm budowania sieci losowej***

Mając dany rozkład sieci docelowej, budujemy ją w następujący sposób: kolejnym  $N$  wierzchołkom przypisujemy stopnie, jakie będą one miały, według rozkładu, idąc od największego stopnia w dół. Następnie, dla pierwszego wierzchołka, który ma mieć stopień  $k_1$ , losujemy  $k_1$  wierzchołków i łączymy go z nimi krawędziami. Wierzchołek numer 2, mający mieć stopień  $k_2$  łączymy z  $k_2$  innymi wierzchołkami, które nie są jeszcze z nim połączone. Krok ten powtarzamy tak długo, aż przejdziemy do ostatniego wierzchołka i połączymy z odpowiednią liczbą pozostałych wierzchołków.

### 4.3.2 Algorytm budowania sieci zachłannej

Kolejnym N wierzchołkom przypisujemy stopnie, jakie będą one miały, według rozkładu, jak poprzednio. Następnie łączymy pierwszy wierzchołek z kolejnymi wierzchołkami na liście. Dzięki temu najpierw łączone krawędziami są wierzchołki o tym samym stopniu, potem o jeden mniejszym... itd. Gdy połączymy pierwszy wierzchołek z odpowiednią liczbą innych wierzchołków, przechodzimy do następnego. Kolejny wierzchołek łączymy znów z wierzchołkami o tym samym stopniu (o ile nie były jeszcze połączone), o stopniu o 1 mniejszym itd. W ten sposób najgęstsze moduły zbudowane są prawie wyłącznie z wierzchołków o największych stopniach.

### 4.3.3 Metodologia tworzenia diagramu obrazującego strukturę sieci pod względem modułowości – badanie procesów dyfuzji.\*

Proces powstania diagramu opisującego strukturę sieci rozpoczynamy od stworzenia macierzy incydencji na podstawie danych o połączeniach w wygenerowanej sieci. Macierz jest generowana na podstawie odpowiednio przygotowanych danych o połączeniach w pliku tekstowym, np.:

$$\begin{array}{cccc} 1 & 1 & 0 & k_1 = 2 \\ 1 & 1 & 0 & k_2 = 2 \\ 0 & 0 & 0 & k_3 = 0 \end{array}$$

Następnie generujemy tzw. „macierz przejścia” za pomocą prostego algorytmu:

```
for i = 1 : rozmiar
  for j = 1 : rozmiar
    if m(i,j) == 1
      wynik(i,j) = 1/stopień(i);
    end
  end
end
end
end
```

---

\* Rozdział napisany na podstawie [13]

Wyliczam również wektory własne na podstawie otrzymanej macierzy. W programach matlabowych posłużyłam się do tego celu prostą funkcją  $eig()$ , która zwraca wartości wektorów własnych i macierz z wartościami własnymi na przekątnej:

Wektory własne  $\rho^{(a)}$ :

$$\begin{array}{ccc} \rho^{(1)} & \rho^{(2)} & \rho^{(3)} \\ 0,525 & 0 & -0,85 \\ -0,85 & 0 & -0,526 \\ 0 & 1 & 0 \end{array}$$

Macierz z wartościami własnymi na przekątnej:

$$\begin{array}{ccc} \lambda^{(1)} = -0,62 & 0 & 0 \\ 0 & \lambda^{(2)} = 0 & 0 \\ 0 & 0 & \lambda^{(3)} = 1,62 \end{array}$$

Na podstawie informacji o wartościach wektorów własnych i macierzy z wartościami własnymi wyliczamy wartości  $c^{(a)}$ , które będą prezentowane na diagramie ze wzoru:

$$c_i^\alpha(t) = \frac{\rho_i^\alpha(t)}{k_i}$$

Według algorytmu postępowania przedstawionego przez Barabasię w artykule „*Diffusion on complex networks: a way to probe their large-scale topological structures*” [2] szukamy takich wartości  $\alpha$ , że  $\lambda^{(\alpha)}$  nie jest równe 1. Poszukujemy tzw. „znaczących” wartości  $\alpha$ , czyli takich, które będą prezentowały na diagramie obraz jak najlepiej odzwierciedlający rzeczywistość i obrazujący modułowość sieci bądź jej brak. Aby je wyznaczyć potrzebna jest sieć pomocnicza „podobna” do naszej sieci wyjściowej z tą samą ilością wierzchołków o podobnym stopniu każdego z nich. Mając rozkład prawdopodobieństwa przyłączenia się nowego wierzchołka do danego wierzchołka sieci o stopniu  $k$  generujemy losowo „podobną” sieć RD (RandOm). Następnie trzeba wyznaczyć Participation Ratio (PR – czyli współczynnik uczestnictwa) dla każdej wartości  $\alpha$  najpierw dla wygenerowanej przez nas sieci, potem dla sieci wygenerowanej losowo. Następnie wybieramy takie dwie wartości  $\alpha$ , że:

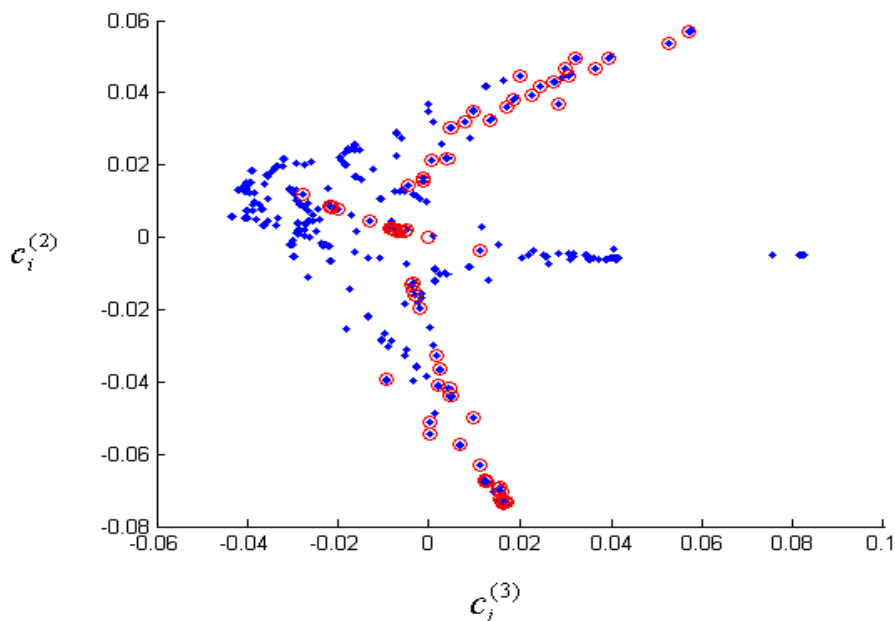
- 1)  $\lambda^{(\alpha)} < 1$  dla sieci wyjściowej
- 2)  $\chi_\alpha$  sieci wyjściowej  $>$   $\chi_\alpha$  sieci wylosowanej
- 1)  $\lambda^{(\alpha)}$  i  $\lambda^{(\alpha)}$  są jak największe (dla sieci wyjściowej)

gdzie:

$$\chi_\alpha = \left[ \sum_{i=1}^N (c_i^{(\alpha)})^4 \right]^{-1}$$

Z analiz Barabasiego

wynika, że odpowiednio stworzony diagram powinien przedstawiać grona w sieci (sieć bliskich znajomości). Na diagramie na niebiesko zaznaczyłam ludzi, którzy tworzą sieć znajomości na podstawie danych ściągniętych programem „grebber.pl” z portalu [www.nasza-klasa.pl](http://www.nasza-klasa.pl). Na rysunku można zaobserwować wyraźnie zarysowującą się „gwiazdzistą” strukturę (punkty układają się w przecinające się proste). Aby potwierdzić słuszność tezy Barabasiego zaznaczyłam na czerwono osoby uczęszczające do tej samej szkoły. Na diagramie obserwujemy teraz co najmniej 3 zbiory czerwonych punktów układające się w proste linie, wokół których skupiają się osoby uczęszczające do tej samej klasy. Na rysunku widać, że już przed oznaczeniem uczniów tych samych szkół można przewidzieć, którzy z nich wzajemnie się znają i uczęszczają do tych samych klas.



Rys. 4.4.3 Diagram przedstawiający strukturę sieci wygenerowanej na podstawie profili użytkowników portalu [nasza-klasa.pl](http://www.nasza-klasa.pl)

#### 4.4 Funkcjonalność programu

Program umożliwia wczytanie sieci i obliczenie jej podstawowych parametrów takich jak ilość wierzchołków oraz współczynnik gronowania. Pozwala też na narysowanie rozkładu sieci oraz analizę topologii na podstawie własności procesów dyfuzji. Ponadto możliwe jest także stworzenie sieci losowych oraz deterministycznej sieci (budowanej zachłannym algorytmem) o takim samym rozkładzie i liczbie wierzchołków. Daje to sposobność do porównania topologii sieci wyjściowej z wygenerowanymi sieciami o takich samych parametrach.

## **4.5 Opis otrzymanych wyników wyjściowych i ich analiza**

Do badań topologii sieci wykorzystam kilka różnych sieci. Większość z nich będzie oparta na danych pobranych z portalu [nasza-klasa.pl](http://nasza-klasa.pl). Dodatkowo wyróżniam sieć Zacharego, która została poddana wnikliwej analizie w artykule Simonsena i innych [3].

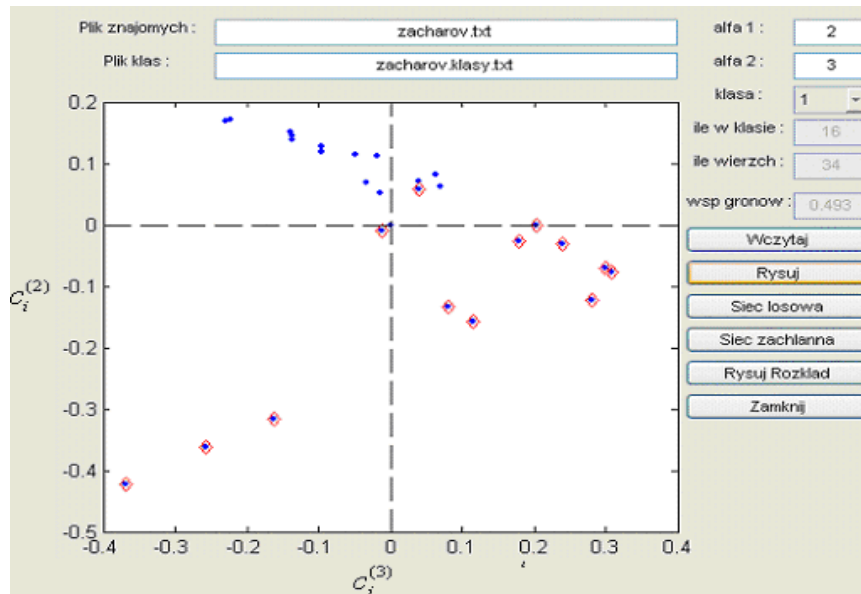
### **4.5.1 Sieć znajomości w sieci Zacharego**

Zrozumienie sensu analizy badań sieci Zacharego wymaga krótkiego wprowadzenia w symulacje sytuacji w realistycznym świecie. Wyobraźmy sobie więc, że istnieje klub sportowy, w którym najważniejsze role odgrywają administrator i trener klubu. Administrator podobnie jak trener mają w członkach drużyny i całej elity klubu zwolenników i przeciwników. Niektóre osoby opowiadały się zarówno za trenerem jak i administratorem, do tego zawierały się pomiędzy członkami przyjaźnie, które mogły znacznie wpływać na podejmowane przez znajomych decyzje.

To jaki wpływ mieli na siebie ludzie nawzajem z klubu Zacharego obrazujemy za pomocą krawędzi w grafie. Poszczególne osoby są prezentowane za pomocą węzłów. Wiedza o przebiegu takich zależności jest nam potrzebna do symulacji sytuacji w przypadku wystąpienia konfliktu pomiędzy trenerem i administratorem. Jak po rozpadzie klubu wyglądałby skład nowopowstałych drużyn?

Mając powyższe informacje przypisujemy ludziom klasy 1 lub 2 odpowiednio dla trenera i administratora. Dzięki programowi możemy przewidzieć na jakie kluby nastąpi rozpad podczas dekompozycji sieci polegającej na usunięciu „słabych więzi” (połączeń

między modułami). Posiadana informacja o rzeczywistym podziale organizacji pozwala nam potwierdzić trafność przewidywań.



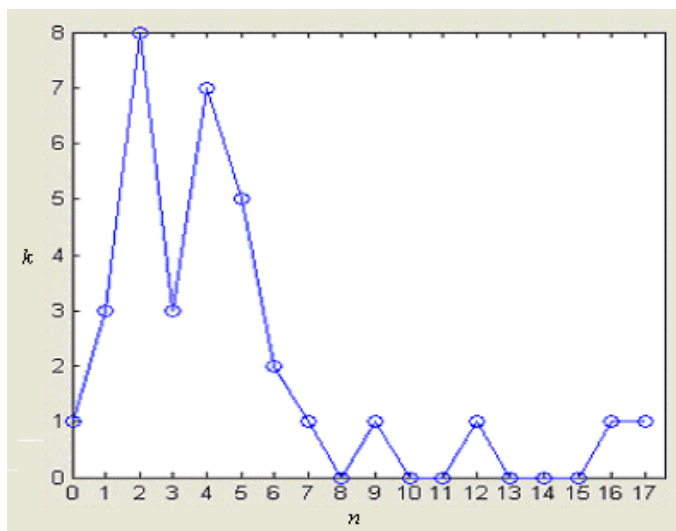
Rys 4.5.1 Diagram przedstawiający strukturę sieci Zacharego

Dla odpowiednio dobranych  $\alpha_1 = 2$  i  $\alpha_2 = 3$  otrzymujemy wykres mówiący nam o przynależności do grup w danej sieci Zacharego. Sieć Zacharego składa się z 34 wierzchołków, mamy dwie klasy. Średni współczynnik gronowania sieci jest równy 0.493. Duży współczynnik gronowania oznacza, że jeśli wierzchołki w sieci potraktujemy jako konkretne osoby, to każda z nich ma dużo sąsiadów, z których wielu zna się na zasadzie „każdy z każdym”. Tam gdzie stopień wierzchołków jest mniejszy na wykresie jest „luźniej”, bo sąsiedzi danego wierzchołka są rzadziej ze sobą połączeni. Łuki zakręcają i zagęszczają się natomiast w tych miejscach, gdzie wierzchołki mają więcej odchodzących od nich krawędzi, bo tam współczynnik gronowania jest duży. Tak wygląda topologia sieci zachłannej.

Na wykresie widoczne są dwie grupy wierzchołków układające się w linie. Przypuszczam, że wierzchołki na jednej linii pokrywają się z grupą trenera, a drugiej z grupą zwolenników administratora. Po zaznaczeniu klasy 1 łatwo zauważyć, że faktycznie wierzchołki z tej grupy należą do klasy administratora. Podobnie dla klasy 2. Jak widać łatwo można zaobserwować własności sieci (zobrazowane na diagramie) nie wiedząc sieci połączeń. Można również przewidzieć co się stanie z siecią po jej ewentualnym rozpadzie, na ile i na jakiej wielkości grupy się najprawdopodobniej rozdzieli np. w razie kłótni pomiędzy administratorem i trenerem klubu.

#### 4.5.1.1 Badanie rozkładu

Na podstawie rozkładu danej sieci (rys.4.5.1.1) możemy stwierdzić, że dużo jest osób znających dwie, cztery bądź pięć osób. Rozkład nie układa w żadną wyraźną strukturę stąd nie możemy jednoznacznie skorzystać z własności rozkładów. Warto jednak zauważyć, że rozkład sieci sugeruje własności sieci bezskalowej - dużo węzłów o mały stopniu i niewielka ilość węzłów o dużym stopniu (występowanie centrów: węzły symbolizujące administratora i trenera). Część rozkładu połączeń między węzłami jest natomiast świadczy o czynniku losowości charakterystycznym dla gęstości połączeń w sieciach losowych (dużo węzłów o zbliżonym, bądź tym samym stopniu). Wielkość średniego współczynnika gronowania (jest duży) i występowanie modułów i centrów sugeruje związek topologii tej sieci z typem sieci bezskalowych, sieć jest jednak zbyt mała aby wysuwać tak daleko idące wnioski.



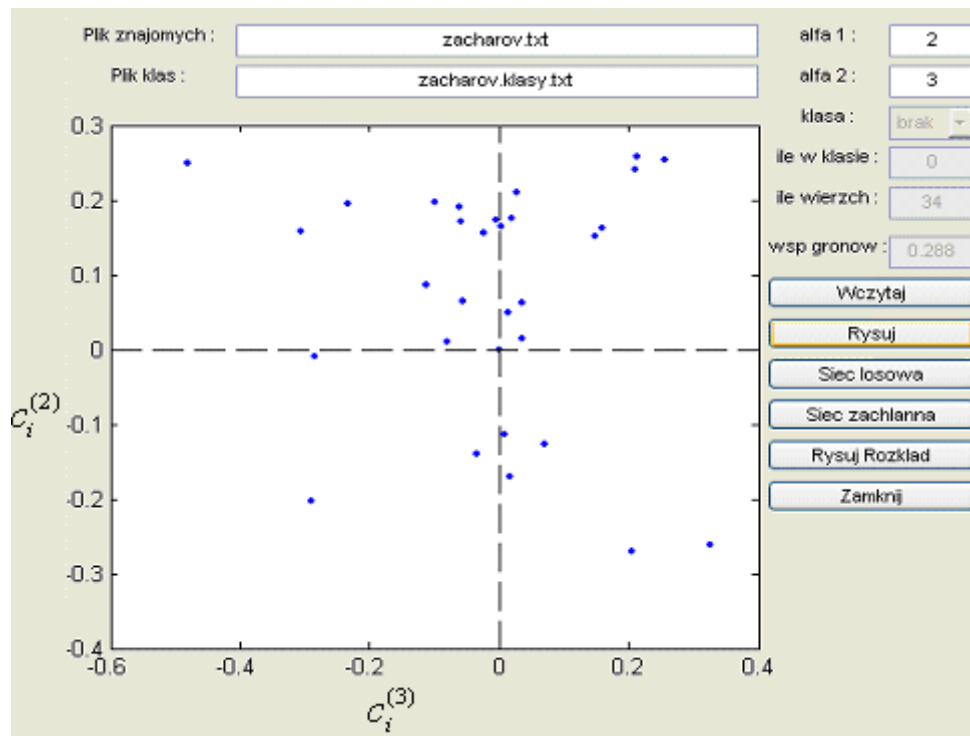
Rys 4.5.1.1 Rozkład gęstości połączeń w sieci Zacharego.

#### 4.5.1.2 Porównanie sieci wyjściowej ze spreparowaną siecią losową

Tworzymy sieć losową o podobnym rozkładzie i tej samej liczbie wierzchołków. Średni współczynnik gronowania wynosi 0.288 i jest znacznie mniejszy od współczynnika gronowania sieci Zacharego. Stąd możemy postawić tezę, że sieć Zacharego jest prawdopodobnie siecią Małego Świata.

Dla tych samych współczynników  $\alpha_1=2$  oraz  $\alpha_2=3$  na wykresie wierzchołki nie układają się w linie proste. Topologia sieci wylosowanej znacznie odbiega od budowy sieci

Zacharego. Pomimo faktu, że rozkład obu sieci jest niemalże identyczny, druga sieć nie posiada wcale wyraźnych modułów. Świadczy to o braku w danej sieci klik, czyli grup osób szczególnie mocno ze sobą powiązanych. Pierwszy nasuwający się wniosek jest prosty: nie można polegać na rozkładzie jako podstawie do identyfikacji struktur sieci.



Rys 4.5.1.2 Diagram sieci Zacharego generowanej algorytmem losowym.

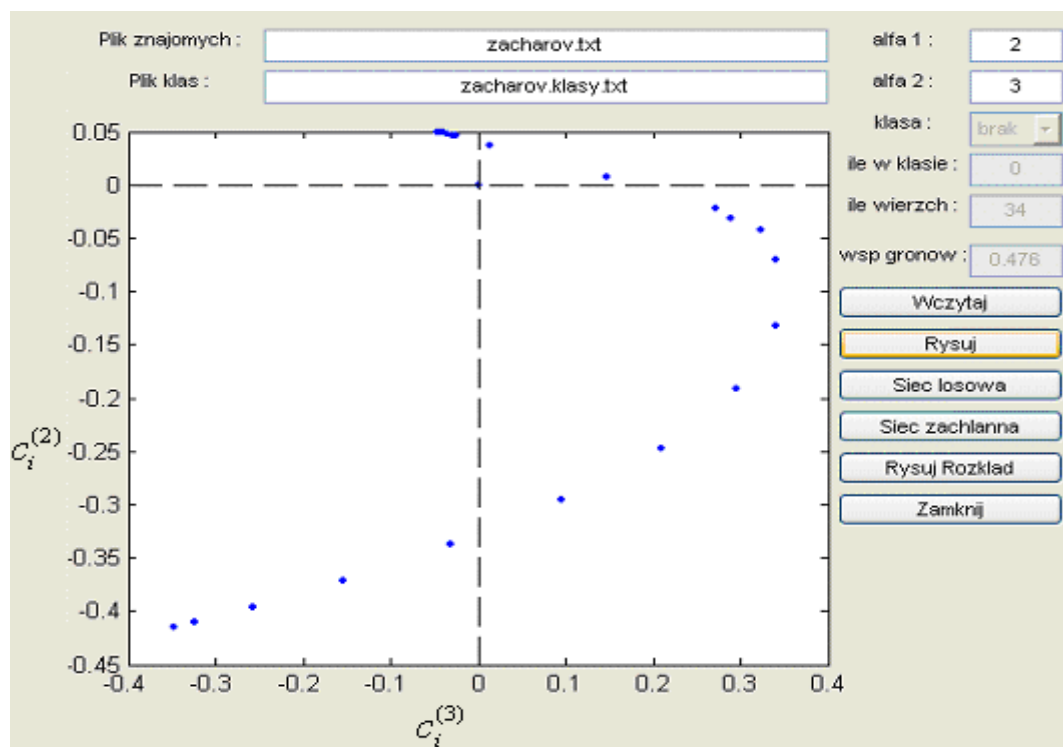
#### 4.2.1.3 Porównanie sieci wyjściowej ze spreparowaną siecią, generowaną algorytmem zachłannym

Generujemy sieć o tym samym rozkładzie algorytmem zachłannym. Dla tych samych  $alfa1 = 2$  oraz  $alfa2 = 3$  na wykresie wierzchołki układają się w spiralną strukturę, która świadczy o zupełnie innej topologii, niż sieć Zacharego. Współczynnik gronowania jest podobny do współczynnika gronowania sieci Zacharego, ale mimo to topologia tej sieci jest zdecydowanie odmienna.

Pierwszym nasuwającym się wnioskiem jest stwierdzenie, że współczynnik gronowania i dany rozkład w oderwaniu od innych parametrów sieci nie mogą zostać użyte do identyfikacji struktury sieci. Dodatkowo na poniższym rysunku widnieją ważne linie proste, co sugeruje strukturę modułową tych sieci. Z łatwością można stwierdzić, na podstawie wcześniej przedstawionej wiedzy, iż struktura wygenerowana siecią zachłanną



odpowiednio przedstawiona graficznie będzie się wyróżniała wyrazistymi modułami i można w niej zaobserwować regularności, które świadczą o niemalże braku czynnika losowości w wybieranych powiązaniach. Regularności są oczywiście związane ze stosowanym algorytmem. Kolejno dobierane powiązania wg algorytmu rozpoczynamy od wierzchołków o jak największych stopniach. W ten sposób tworzy się klika.



Rys 4.5.1.3 Diagram sieci Zacharego generowanej algorytmem zachłannym.

#### 4.5.2 Sieć znajomości Nasza Klasa 1

Sieć wygenerowana, którą w tym podrozdziale będę nazywać „Nasza Klasa 1”, jest siecią utworzoną z profili użytkowników portalu [nasza-klasa.pl](http://nasza-klasa.pl). Użytkownicy od których program rozpoczyna pobieranie danych zostali wybrani z grupy wspólnych znajomych. Program pobrał i połączył więzami 2511 użytkowników, którzy byli znajomymi, lub mieli wspólnych znajomych. Na podstawie przedstawionego poniżej diagramu określe parę własności struktury wygenerowanej sieci.

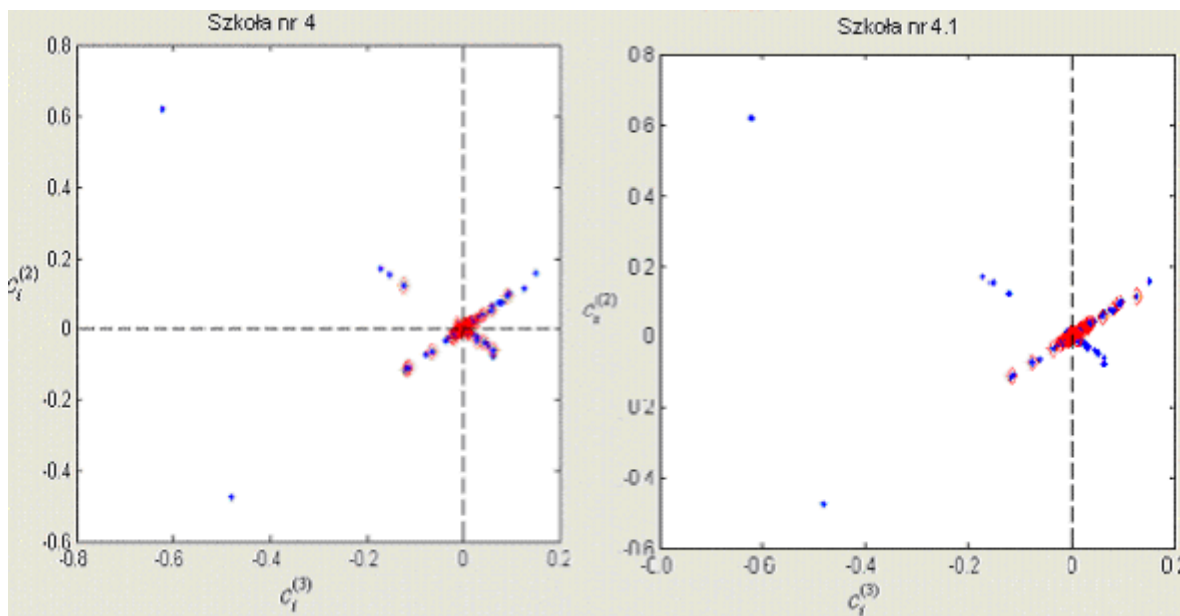
Program po stworzeniu sieci wyświetla również wartość średniego współczynnika gronowania, która służy do oszacowania, ilu sąsiadów danego wierzchołka jest połączonych „każdy z każdym”. Jest to iloraz liczby krawędzi pomiędzy sąsiadami danego wierzchołka,

przez ilość krawędzi, jaki miałby graf pełny składający się z tych sąsiadów. Średni współczynnik gronowania ma dość niską wartość 0,07 jednakże w przypadku tak wielkiej sieci (im większa sieć tym większa ilość możliwych połączeń) jego wielkość jest i tak dość okazała.

O występowaniu wyraźnych modułów w sieci (grup dobrze znających się znajomych, osób uczęszczających do jednej klasy) świadczy w przejrzysty sposób diagram stworzony przez program „Analiza sieci”. Zarysowują się na nim dwie grupy punktów układające się w przecinające się wzajemnie linie proste. Jak już wspomniałam proste obrazują moduły (tu klasy, szkoły lub grupy znajomych). Dodatkowo na prostych można wyróżnić gęściej skupione punkty, czyli osoby pomiędzy którymi wytworzyły się bardzo bliskie relacje. Osoby prezentowane jako gęsto skupione w grupie punkty znają się wzajemnie w relacji niemalże „każdy z każdym”.

Ponieważ program grabber.pl pobiera najpierw profile osób dodanych jako pierwsze do listy znajomych (programowi muszą niestety zostać nadane pewne ograniczenia na ilość znajomych których pobierze dla danego profilu ze względu na zbyt dużą rozległość sieci i ograniczone możliwości Matlaba oraz mocy obliczeniowej mojego komputera). Łatwo wywnioskować na podstawie diagramu, iż są to najczęściej osoby pochodzące z tej samej co dany użytkownik szkoły lub klasy. Aby potwierdzić fakt przynależności osób uczęszczających do tej samej klasy prezentowanych jako punkty skupione wokół linii prostych na wykresie zaznaczyłam na czerwono węzły, które symbolicznie przedstawiają uczniów jednej szkoły (w tym przypadku nr 9). Gwieździsta struktura układu węzłów obrazujących uczniów uczęszczających do tej samej szkoły pokrywająca się z liniami prostymi na wykresie w miejscach gdzie punkty tworzą osadzone blisko siebie grupki potwierdza postawioną na początku przeze mnie tezę. Można również wstępnie stwierdzić, że wybrani użytkownicy (w liczbie 1021) należą przede wszystkim do dwóch klas w danej szkole nr 4 (mogą to być klasy, których podział jest traktowany rocznikowo) lub szkoła jest podzielona na dwa różne profile, albo też umieszczona jest w dwóch różnych budynkach.

Aby się przekonać jak powyższe tezy odwołują się do rzeczywistości zaznaczyłam na drugim wykresie klasę (lub omawianą szkołę, ale o konkretnym profilu), która wyraźnie jest podgrupą wcześniej zaznaczonej a jej uczniowie są osadzeni na jednej linii prostej. Grupa osób uczęszczająca do szkoły nr 4.1 jest prezentowana na rysunku poniżej. Nazwałam ją szkołą gdyż liczba uczniów jest na tyle duża (521 węzłów) aby klasyfikować tę grupę na tym poziomie. Oczywiście wyniki analiz można interpretować różnie, mogą to być również studenci jednej z dwóch specjalności na danym kierunku studiów itp.



Rys 4.5.2 Diagramy z zaznaczonymi na czerwono uczniami wybranych szkół

#### 4.5.2.1 Badanie rozkładu

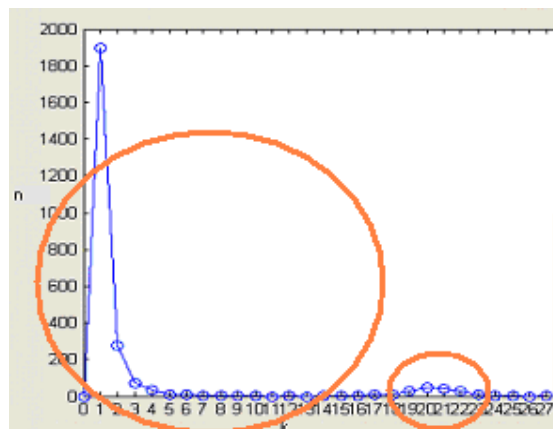
Na rysunku 4.5.2.1 obserwujemy znacznie przeważającą liczbę węzłów o niewielkich stopniach. Z takim zjawiskiem będziemy mieli do czynienia w przypadku rozkładu sieci połączeń tworzonych na podstawie danych z portalu [nasza-klasa.pl](http://nasza-klasa.pl) ponieważ generowanie sieci tworzonej przez program musi być w pewnym momencie przerywane. Stąd pozostaje wiele węzłów o stopniu 1. Ze względu na ten fakt i wiążące się z tym zagrożenie wkładnięcia się zakłamań w wyniki badań często nie biorę ich pod uwagę podczas analizy.

Jak już wcześniej wspominałam model sieci losowych opisuje się rozkładem  $P(k)$ , gdzie  $k$  jest stopniem wierzchołka a większość węzłów ma w przybliżeniu tę samą ilość odchodzących od niego krawędzi. Wierzchołki posiadające stopień różny od  $k$  są rzadkie. Im większe odchylenie od  $k$ , tym mniej takich wierzchołków.

Analizowana sieć *Nasza Klasa 1* posiada właściwości sieci losowej gdy weźmiemy pod uwagę wierzchołki stopnia 2 (jak już wspomniałam wcześniej węzły stopnia 1 pomijam). Jest ich zdecydowanie więcej niż wierzchołków innego stopnia. Jednocześnie węzły o stopniu bliskim 2 – stopnia 3 i 4 - są najbardziej zbliżone pod względem wielkości częstości występowania. Wierzchołki posiadające stopień różny od 2 są rzadkie i wraz z wzrostem odchylenia od  $k = 2$  zmniejsza się ilość wystąpień tych wierzchołków. Jednakże taka analiza

rozkładu jest jego dużym uproszczeniem i może być poprawna jedynie względem części z przedstawionej sieci.

Sytuacja zmienia się gdy pod lupę weźmiemy inny fragment rozkładu oscylujący w okolicach stopnia węzła  $k = 20$ . Na podstawie rysunku 4.5.2.1 można stwierdzić, że rozkład ilości połączeń odchodzących od węzłów jest postacią zbliżony kształtem do rozkładu Poissona. Określając rozkład jako podobny nie biorę pod uwagę węzłów o małych stopniach. W sieci jest dużo węzłów o stopniu 20, ale także o zbliżonych do tej liczby stopniach: 19, 21, 22. Jest to efekt ograniczenia programu do pobierania profili nowych znajomych w określonej, maksymalnej liczbie ze względu na posiadanie przez wielu użytkowników portalu grupy znajomych w liczbie przekraczającej 100. Danych znajomych ponad pewną, ustaloną w parametrach wejściowych liczbę nie pobieram przez program gdyż stworzone w ten sposób sieci byłyby bardzo rozłożyste. Można więc wziąć pod uwagę fakt, że sieci tworzone przez program nie oddają dokładnie rzeczywistej struktury powiązań w portalu a węzły o stopniu 20 są w rzeczywistości węzłami o stopniu znacznie przewyższającym tą liczbę, np. 90, 203.

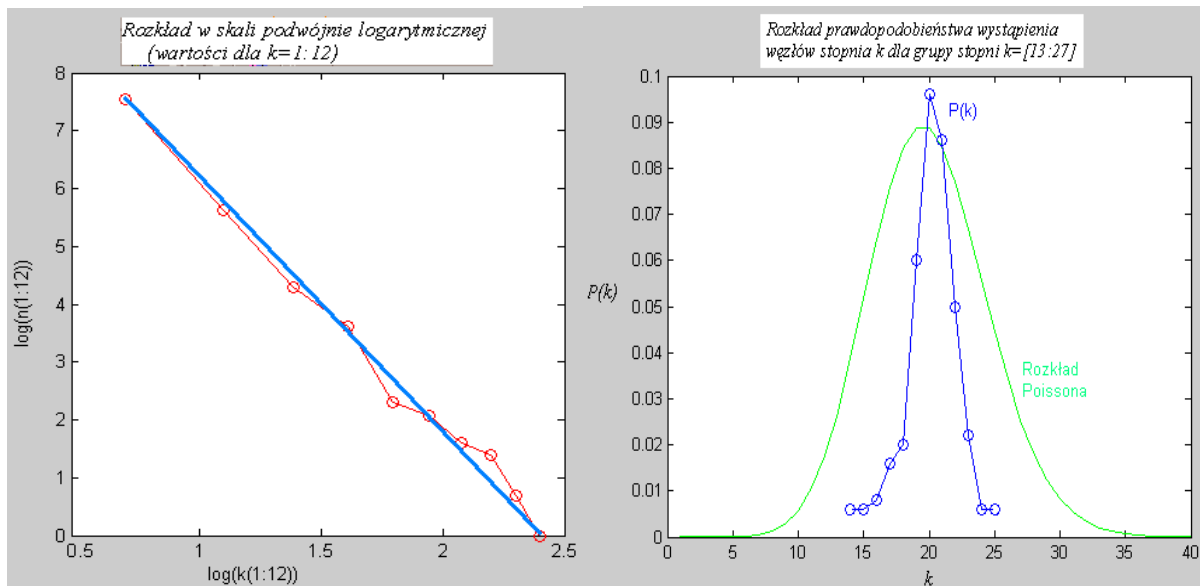


Rys. 4.5.2.1 Rozkład liczby połączeń w sieci nasza Klasa 1

Natomiast gdy spojrzymy na całą ogólną strukturę sieci, łącznie z wierzchołkami o niewielkim stopniu oraz pominiemy fakt istnienia zwiększonej ilości węzłów o stopniu w okolicach liczby 20 zauważymy podobieństwo rozkładu ilości połączeń do kształtu funkcji potęgowej (charakterystyczny dla sieci bezskalowej). W skali podwójnie logarytmicznej rozkład niemal pokrywa się z funkcją prostą, co potwierdza te przypuszczenia. Aby bliżej zaobserwować mieszanie się tych dwóch własności rozkładów Poissona (w obrębie wierzchołków o dużym stopniu) i potęgowej (stosowanego w skali ogólnego spojrzenia na rozkład węzłów sieci) podzieliłam rozkład ilości połączeń w sieci na dwie części i

sprawdziłam czy są jemu bliskie kształty tych dwóch rozkładów. Rozkład Poissona

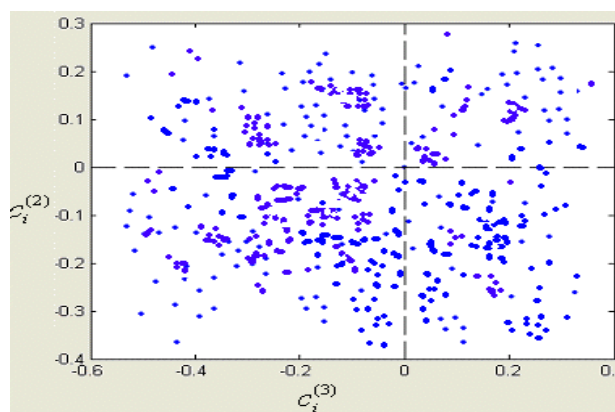
$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$  na wykresie jest narysowany dla  $\lambda = 20$ , gdzie  $k = 0, 1, 2 \dots 40$ .



Rys 4.5.2.2. a) rozkład stopni węzłów o stopniach od 1 do 12 w skali podwójnie logarytmicznej, b) rozkład prawdopodobieństwa wystąpienia węzłów danego stopnia w grupie węzłów o stopniach 13-17.

#### 4.5.2.2 Porównanie sieci wyjściowej ze spreparowaną siecią losową

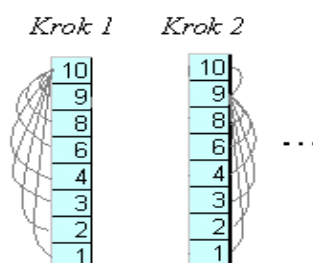
Na podstawie rozkładu sieci wygenerowanej za pomocą algorytmu opierającego się o losowy wybór powiązań generujemy nową sieć a na jej podstawie diagram. Między krawędziami na rysunku 4.5.2.2 punkty nie układają się w linie bądź gwiazdziste struktury – punkty(węzły) są rozrzucone losowo. Na rysunku 4.5.2.2 widać, że sieć generowana algorytmem losowym zyskuje właściwości sieci przypadkowej. Charakteryzuje ją brak wyraźnych odznaczających się skupisk węzłów symbolizujących centra i układania się punktów w linie proste reprezentujące moduły. Średni współczynnik gronowania sieci losowej jest mniejszy od średniego współczynnika klasteryzacji sieci Małego Świata, której cechy posiada sieć przedstawiona na rysunku 4.5.1.



Rys 4.5.2.2 Diagram sieci generowanej algorytmem opartym o losowy wybór powiązań

#### 4.5.2.3 Porównanie sieci wyjściowej ze spreparowaną siecią, generowaną algorytmem zachłannym

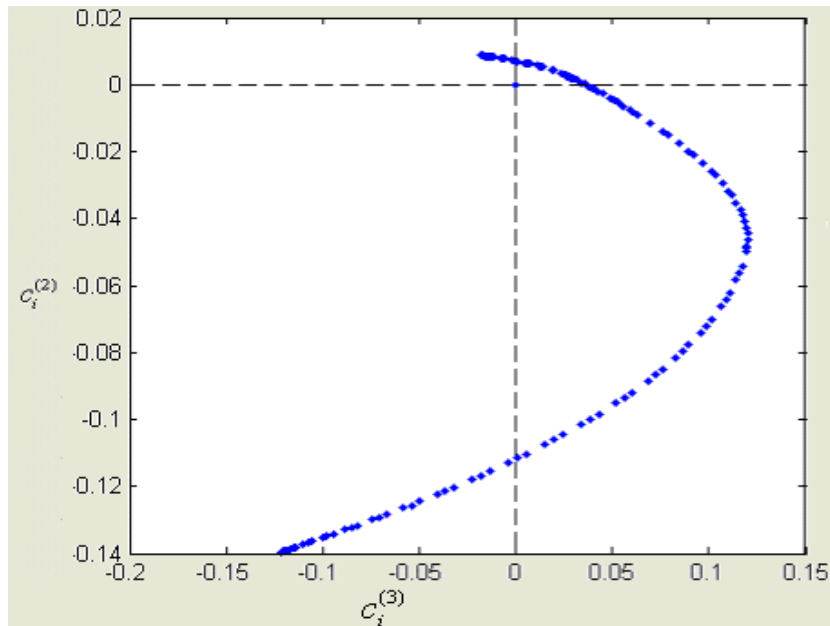
Algorytm tworzenia sieci metodą zachłanną pozwala już na wstępie stwierdzić, iż generowana za pomocą takiego postępowania sieć będzie się wyróżniała dużą regularnością



Rys 4.5.2.3 a) Algorytm generowania sieci zachłannej na podstawie rozkładu gęstości sieci Nasza Klasa 1. Kolejność wyboru następnego połączenia jest uwarunkowana przez stopień jaki posiada dany wierzchołek.

Obserwacja diagramu potwierdza wcześniejsze oczekiwania. Jego struktura sugeruje występowanie w nowopowstałej sieci modułów. Jest to wyraźny łuk, który przy odrobieniu wyobraźni możemy potraktować jako dwie linie łączące się w pewnym punkcie. Dlaczego wykres jest tak regularny i przybiera taką postać? Otóż podczas generowania takiej sieci bierzemy najpierw wszystkie węzły, które mają największy stopień i je łączymy z węzłami z coraz to mniejszymi stopniami (w tablicy pomocniczej rejestrujemy każde połączenie i śledzimy ciągle zmieniającą się liczbę dostępnych wolnych powiązań). Jeśli brakuje kolejnych węzłów o różnym malejącym stopniu zaczynamy od początku ponownie od największego, który ma dostępne wolne „gałęzie”. Węzły o największych stopniach są więc powiązane ściślej ze sobą tworząc zwarty moduł. Część sieci prezentowanej u góry rysunku przez osadzone blisko siebie punkty jest modułem o dużej gęstości.

Na końcu pozostają węzły o najmniejszym stopniu z pozostałych, które nie przyłączyły się jeszcze do żadnej sieci i szukają węzłów, które mają największy dostępny stopień (algorytm zaczyna iść „od dołu”). Taka zmiana postępowania ma odzwierciedlenie na wykresie w postaci łuku. Tworzony jest drugi moduł, którego gęstość jest mniejsza (większe przestrzenie w dolnej części diagramu). Oba moduły nie są jednak od siebie odcięte tylko łączą się wzajemnie.



Rys 4.5.2.3 b) Diagram sieci generowanej algorytmem opartym o wybór powiązań między węzłami algorytmem zachłannym na podstawie rozkładu sieci Nasza Klasa 1

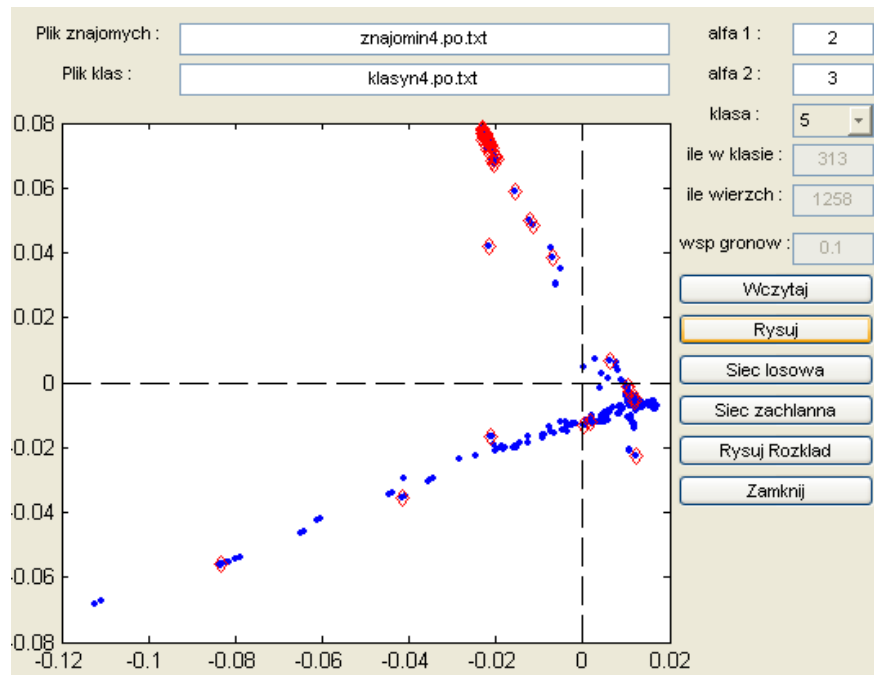
#### 4.5.3 Sieć znajomości Nasza Klasa 2

Kolejna sieć „Nasza Klasa 2”, którą wygenerowałam za pomocą programu grabber.pl różni się od poprzedniej jednakże obserwacje jej prowadzą do podobnych wniosków. Program połączył więzami 1258 użytkowników portalu sugerując się obecnością na liście znajomych. Wartość średniego współczynnika gronowania wynosi 0,1 i jest wartością wyższą od prezentowanej dla poprzedniej sieci. Wiemy zatem, że powiązań w tej sieci jest więcej (odznacza się większą gęstością połączeń).

Na diagramie zarysowują się na dwie grupy punktów układające się w przecinające się wzajemnie linie proste. Na prostych można wyróżnić gęściej skupione wokół siebie punkty czyli osoby pomiędzy, którymi wytworzyły się bliższe relacje (każdy z każdym).

Warto również wspomnieć, że osoby, od których rozpoczynamy budowanie sieci (zwani dalej inicjatorami) są przypadkowo wybrane i podział na dwie grupy punktów układających się w linie również to obrazują. Podobnie jak poprzednio aby potwierdzić fakt przynależności osób uczęszczających do tej samej klasy prezentowanych jako punkty skupione wokół linii prostych na wykresie zaznaczyłam na czerwono węzły określone jako przynależące do jednej szkoły (nr 4). Na tej podstawie można stwierdzić, że wyróżnieni czerwonym kolorem użytkownicy (w ilości 313) należą do jednej klasy i są głównie

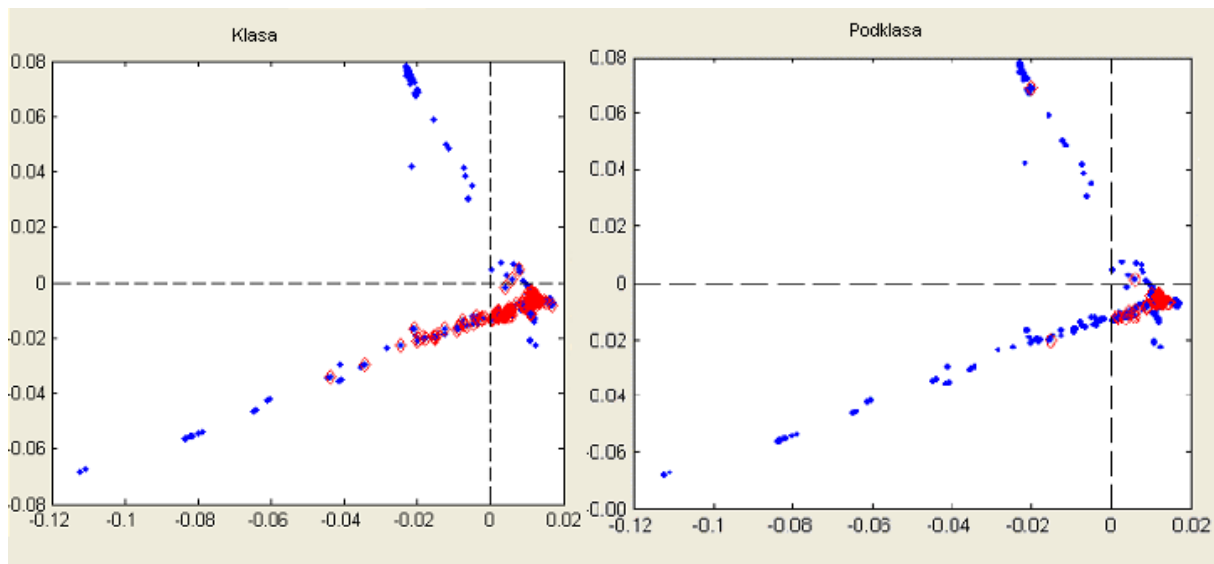
znajomymi jednego z wybranych losowo użytkowników portalu. Zdecydowana większość punktów leży na linii prostej, tylko nieliczni znajomi należą do różnych szkół, klas.



Rys 4.5.3 a) Diagram sieci wygenerowanej na podstawie danych zebranych z portalu *nasza-klasa.pl* - sieć Nasza Klasa2

Na diagramach zaznaczyłam również inne obszary, które symbolicznie nazwałam klasami, bowiem nie zawsze nimi są w rzeczywistości i można je różnie interpretować. Warto również wspomnieć, że połączenie linii i zagęszczenie liczby punktów w tym miejscu oznacza, że grupa znajomych jest wyjątkowo ściśle powiązana ze sobą wzajemnie a jako punkt wspólny schodzących się modułów może być przykładem środowiska znajomych uczęszczających do dwóch klas jednocześnie (klasa symbolizuje tu szkoły, kierunki studiów, wspólnych znajomych – być może dalekich – użytkowników traktowanych jako inicjatorów generowanych sieci). Zaznaczona powyżej klasa obejmuje znajomych też znajdujących się na drugiej linii prostej, których powiązania między sobą są bardzo odległe. Jednakże jako węzły, z których generowane będą następne połączenia pozwalają przewidzieć skupienie się ostatecznie dwóch linii w punkcie gromadzącym wokół siebie wspólnych, mocno ze sobą powiązanych znajomych (zjawisko, które można nazwać zapętleniem sieci).





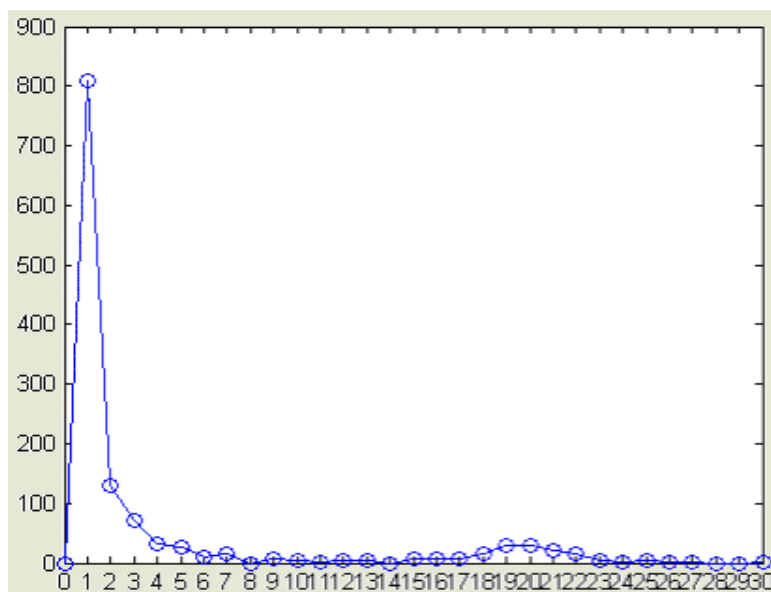
Rys 4.5.3 b) Diagram sieci wygenerowanej na podstawie danych zebranych z portalu *nasza-klasa.pl* – przykład klasy i podklasy dla sieci *Nasza Klasa 2*

Na rysunkach widać wyraźnie sieci jako zbiór modułów i podmodułów, szkół i klas, grup i podgrup znajomych.

#### 4.5.3.1 Badanie rozkładu

Podobnie jak w przypadku wcześniej analizowanej sieci rozkład gęstości sieci ma podobny kształt i wyróżnia się podobnymi własnościami. Sugeruje strukturę sieci na pograniczu sieci Małego Świata, losowej, ale przede wszystkim bezskalowej. Takie sieci często mają względnie małą liczbę węzłów o wysokim stopniu (centra, huby) w stosunku do ilości węzłów o znacznie niższych stopniach wierzchołków (dużo węzłów o stopniach 1, 2, 3, 4, 5).

Podobnie jak poprzednio, można zaobserwować większą ilość osób o liczbie znajomych bliskiej 20. Kształt rozkładu przybiera w tym momencie formę rozkładu Poissona (sieci losowe). Ponieważ ta anomalia jest bardziej efektem świadomie wprowadzonego ograniczenia na stopień węzłów, niż przypadku będą tę sieć zaliczać do grona sieci bezskalowych.

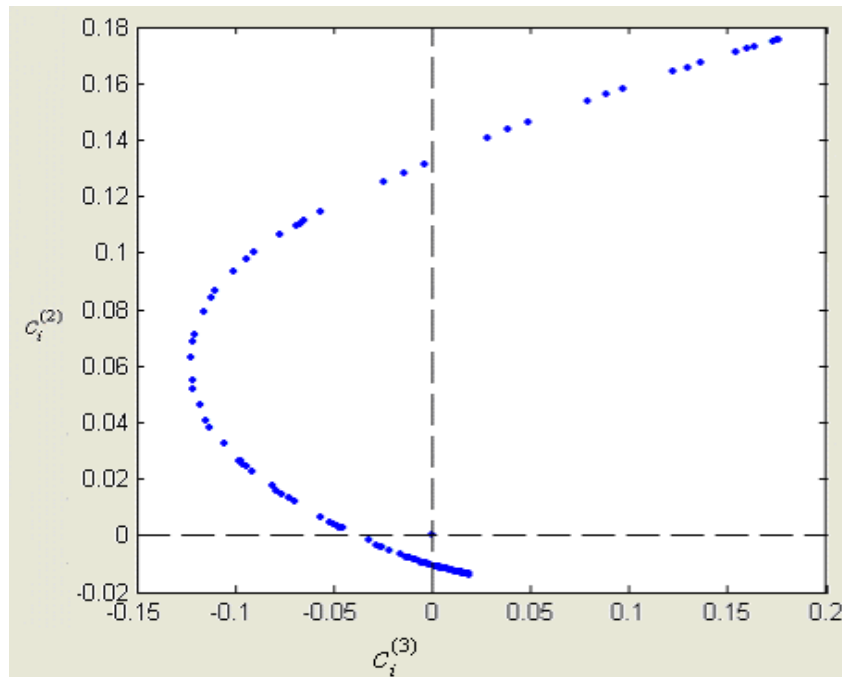


Rys 4.5.3.1 Rozkład stopni węzłów sieci Nasza Klasa 2

Na bazie zebranych danych wnioskuję, że grupa znajomych określonych jako Nasza Klasa 1 (podobnie jak Nasza Klasa 2) tworzy sieć bezskalową. Ta niewielka społeczność jest zbiorowiskiem osób, z których większość ma kilku dobrych bliskich znajomych. Liczba osób wyjątkowo popularnych jest znacznie mniejsza. Dodatkowo jeśli osoba znajdująca się w tej sieci będzie miała ponad 8 znajomych to z dużym prawdopodobieństwem będzie ich posiadała w liczbie 20 os.

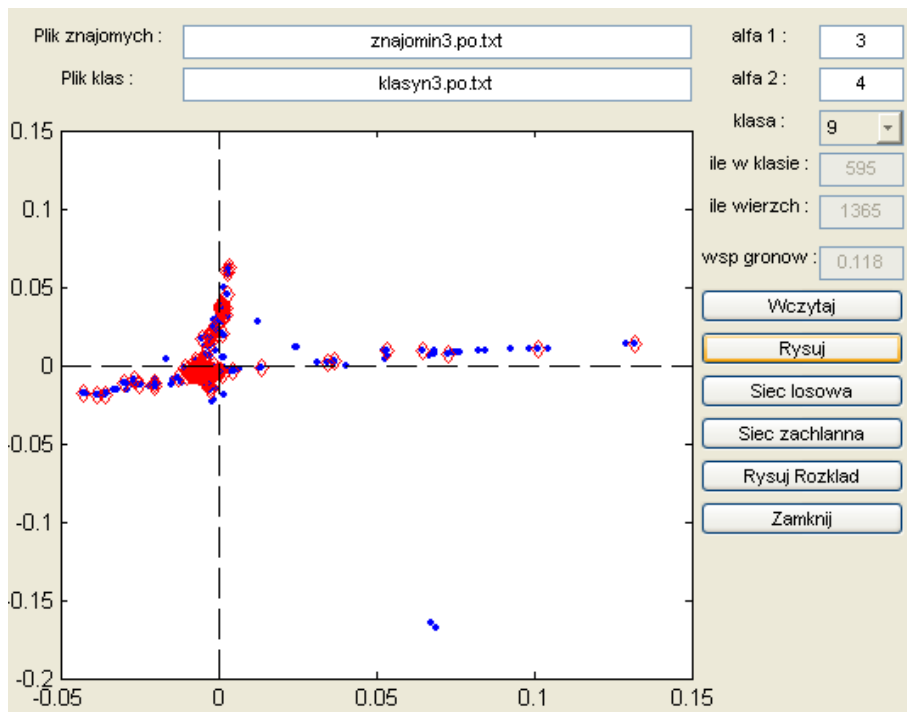
#### 4.5.3.2 Porównanie sieci wyjściowej ze spreparowaną siecią, generowaną algorytmem zachłannym i losowym

Sieć generowana algorytmem zachłannym nie zaskakuje nas swoją strukturą, jest bardzo podobna do struktury sieci *Nasza Klasa 1*. Algorytm postępowania jest bowiem ten sam, zmienia się jedynie liczba węzłów i ilość więzów pomiędzy nimi. Diagram jest regularny i przybiera postać łuku. Podobnie jak poprzednio mamy do czynienia z zagęszczeniem liczby połączeń, tym razem jednak u dołu diagramu. Na diagramie można zaobserwować dwa główne moduły.

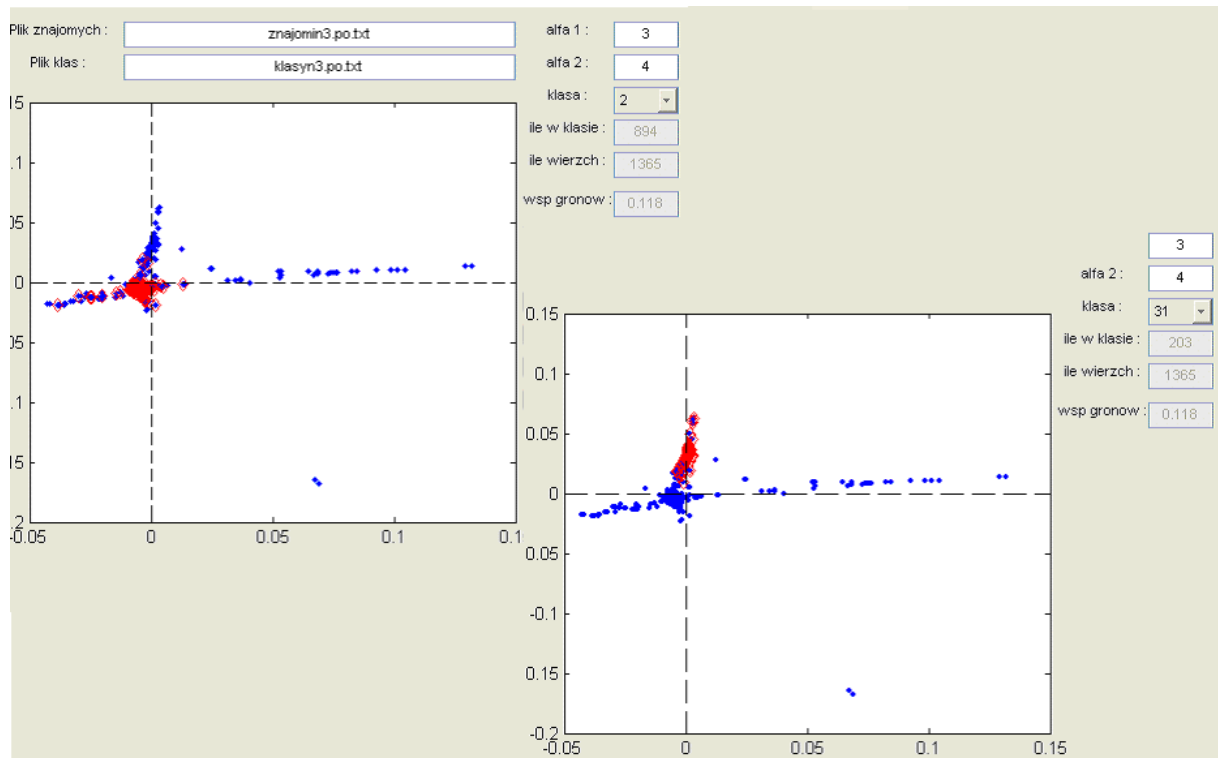


Rys 4.5.3.2 Diagram przedstawiający strukturę sieci generowanej algorytmem zachłannym na podstawie danego rozkładu.

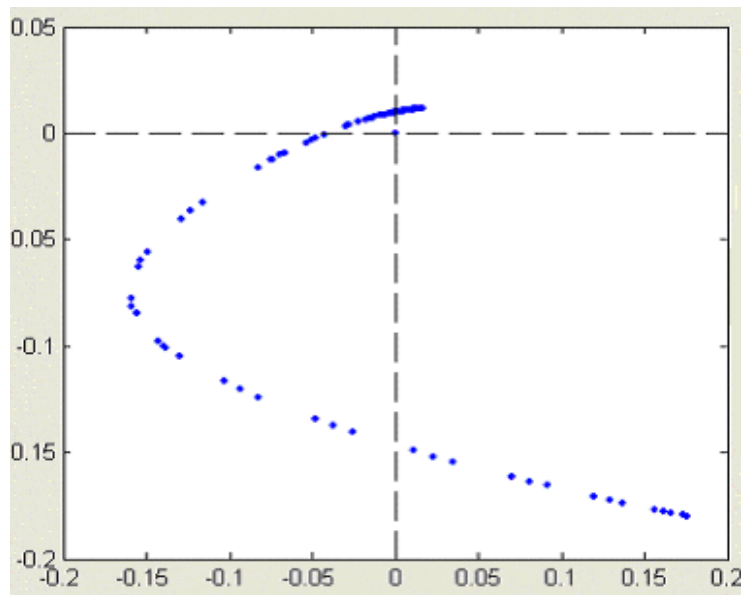
#### 4.5.4 Sieć znajomości Nasza Klasa 3



Rys. 4.5.4.1 Diagram prezentujący wybrana klasę w sieci

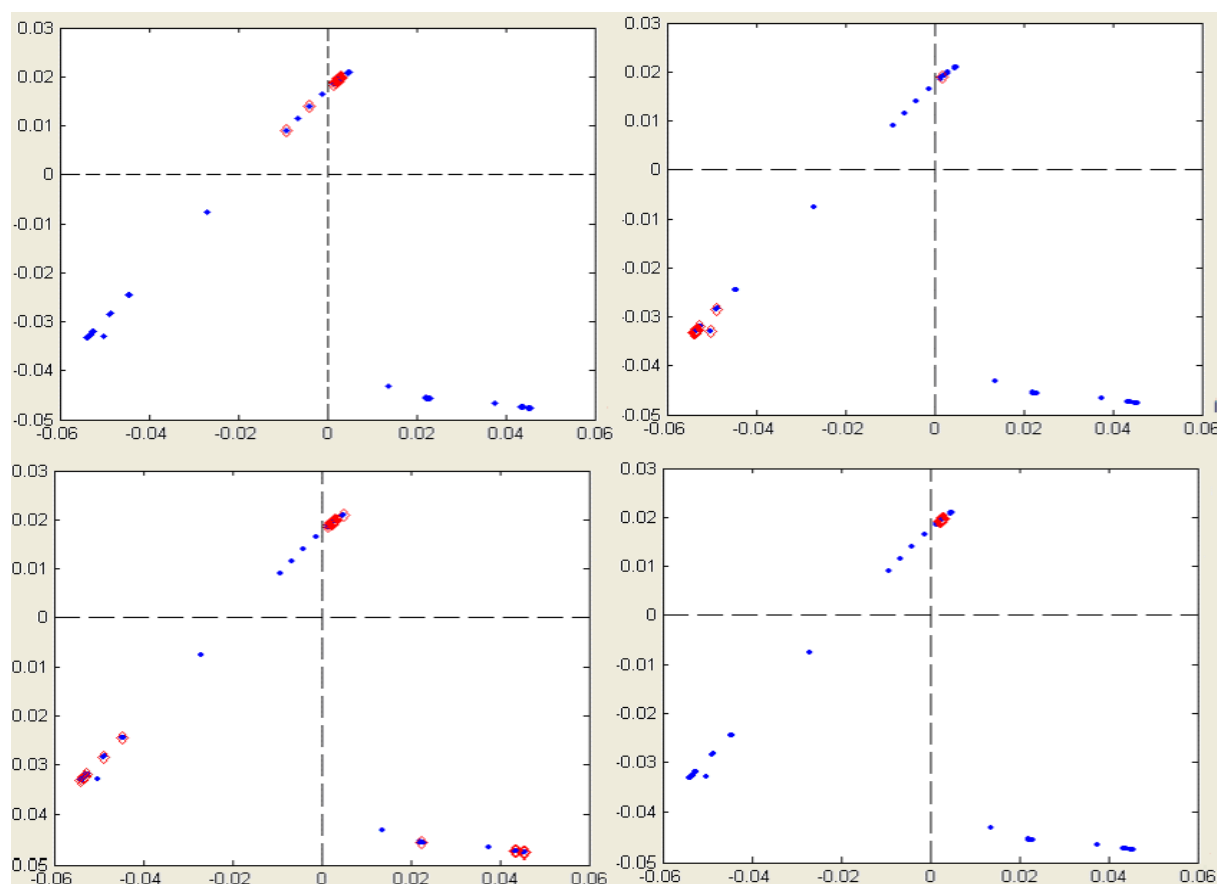


*Rys. 4.5.4.2 Rysunek prezentujący wybrane podklasy wygenerowanej sieci*



*Rys. 4.5.4.3 Diagram sieci prezentujący strukturę sieci generowanej algorytmem zachłannym na podstawie rozkładu gęstości połączeń sieci Nasza Klasa 3*

#### 4.5.5 Sieć znajomości i zaznaczone w niej moduły(podgrupy) - Nasza Klasa 4



Rys. 4.5.5 Diagramy sieci prezentujące strukturę sieci Nasza Klasa 4 jako zespół modułów i podmodułów. Sieć składa się z 1274 węzłów i posiada średni współczynnik gronowania: 0,11. Na kolejnych diagramach zaznaczone są różne „klasy”

## 5. Zastosowanie wiedzy o modułach w sieci

Badanie sieci pod względem ich modułowości ma zastosowanie w wielu dziedzinach. Znając informacje o budowie sieci i zasady rządzące światem możemy stać się świadkami interesujących odkryć i twórcami nowych reguł, które pomogą przewidzieć zachowania z pozoru wyglądające na nie nadające się do ogarnięcia przez żadną prawidłowość. Przy wykorzystywaniu wiedzy o sieciach powstało przykładowo twierdzenie o „sile słabych więzi” (strength of weak ties), sformułowane przez Marka Granovettera.

Wskazuje ono na to, że jeśli dwie osoby połączone są mocną relacją, na przykład przyjaźnią, to grona ich znajomych w dużym stopniu się pokrywają, podczas gdy osoby połączone słabymi relacjami będą miały różnych znajomych – zarówno bliskich jak i tych bardziej odległych. W konsekwencji sieci osób połączonych mocnymi więziami są bardzo gęste i zamknięte, podczas gdy obecność słabych więzi powoduje rozluźnienie struktury

sieciowej. Przykładem ilustrującym znaczenie tego twierdzenia jest poszukiwanie pracy. Okazuje się, że osobami, do których powinniśmy zwracać się z prośbą o pomoc w znalezieniu zajęcia nie są wcale najbliżsi, ale osoby, które są nam znane słabiej. Przyjaciele nie będą w stanie nam pomóc, ponieważ znają te same osoby, do których najprawdopodobniej i tak sami byśmy się zwrócili. Bliscy znajomi nie są w stanie powiedzieć niczego, czego byśmy już sami nie wiedzieli. Co innego dalsi znajomi – ich krąg znajomości w mniejszym stopniu pokrywa się z naszym, dlatego za ich pośrednictwem łatwiej dowiedzieć się o szansach na etat [5].

Ten przykład pokazuje, że jakość kapitału społecznego zależy od strukturalnych cech sieci społecznej. Jeśli miarą jakości kapitału społecznego jest jego zdolność do dostarczania nam nowych informacji o szansach na korzystne działanie, to taki kapitał będzie generowany przez rzadką sieć kontaktów opartą o słabe więzi. Dodatkowo uzasadnia to sens wnikliwej analizy świata w ujęciu sieciowym.

Reguła słabych więzi została również sformalizowana w języku teorii tendencji. Elementy nielosowe w sieci obejmują bowiem różnego rodzaju tendencje. Można do nich zaliczyć efekty czysto strukturalne, związane z właściwościami samych relacji społecznych, oraz preferencje jednostek, ich indywidualne skłonności lub upodobania, na przykład, skłonność do nawiązywania relacji z osobami podobnymi do siebie. W obrębie sieci teoria tendencji pozwala na przedstawienie warunków, w jakich społeczeństwo charakteryzować się będzie możliwie największą spójnością, w jakich zaś będzie ono wysoce podatne na rozpad, dekompozycję [1].

## 6. Podsumowanie

Z moich badań wynika, że za pomocą przedstawionych w artykule metod obliczeniowych możliwe jest określenie topologii zadanej sieci, tzn. wyodrębnienie modułów, czyli gęsto skoncentrowanych węzłów, podmodułów oraz centrów. Informacje o modułach w sieci złożonej można wykorzystać, pozwalają bowiem :

- przewidzieć, na jakie rozdzielne sieci rozpadnie się sieć wyjściowa, jeśli sieć zostanie zdekomponowana poprzez usunięcie „słabych” więzów pomiędzy modułami.
- rozpoznać do jakich klas/grup znajomych należały zadane osoby, mając jedyne informacje o znajomościach pomiędzy poszczególnymi osobami.
- określić jak zorganizowana jest firma (jakie są zespoły, lub jak rozmieszczeni są pracownicy w biurach) bez wiedzy o składzie zespołów mając jedynie informacje o tym kto się z kim kontaktuje podczas pracy.

Program „Analiza sieci” pomaga również w dokładnej analizie sieci pod wieloma względami, dzięki swej przejrzystości. Często gmatwanina połączeń pomiędzy węzłami w dużej sieci niewiele nam mówi. Przedstawiona w przemyślany sposób i odpowiednio zanalizowana może być źródłem zaskakujących i przydatnych informacji. Struktura sieci wymaga dokładnej analizy. Ja skupiłam się na rozpatrzeniu jej pod względem modułowości i identyfikowaniu ze znanymi strukturami sieciowymi, co w dużym stopniu udało mi się zrobić dzięki zaprojektowanemu programowi.

Starłam się również pokazać jak można przewidzieć z dużym prawdopodobieństwem przynależność do grup znajomych, klas, specjalistów z tej samej dziedziny wyłącznie na podstawie odpowiednio sformułowanych sieci. Jak widać na diagramach w poprzednich rozdziałach w wielu przypadkach przewidywania pokrywały się z realnymi danymi. Dodatkowo zanalizowałam sieci pod względem ilości połączeń między węzłami co pomogło w zakwalifikowaniu ich do konkretnej grupy sieci. Wprowadziłam również modyfikację do sieci i zmieniałam algorytmy ich budowy aby zbadać wpływ tych czynników na zmianę własności sieci. Okazało się, że zmiana w algorytmie budowania sieci, nawet przy identycznym rozkładzie gęstości połączeń ma decydujące znaczenie przy kwalifikacji pod względem rodzaju sieci i ma wpływ na sposób klasteryzacji w jej obrębie. Żaden z czynników kwalifikujących sieć do danej grupy (rozkład połączeń, współczynnik gronowania...) nie może samodzielnie jednoznacznie o niej przesądzać.

Praca miała pokazać jak można analizować topologię sieci i identyfikować ją ze zdefiniowanymi już przez naukowców rodzajami sieci złożonych żeby przy opisie skorzystać z ich własności. Dodatkowo starałam się przedstawić niektóre z wielu interpretacji takich sieci aby uzmysłowić użyteczność takich badań, i interdyscyplinarność nauki o sieciach. Sieci złożone są bardzo zróżnicowane. Podstawowym czynnikiem różnicującym jest zróżnicowanie węzłów. To właśnie przede wszystkim od rodzaju węzłów zależy charakter sieci. Węzłami w sieci mogą być bardzo różnorodne obiekty np. liczba rzeczywista, osoba, wektor. Kolejnym zróżnicowaniem występującym w sieciach złożonych jest zróżnicowanie połączeń. Ma ono decydujący wpływ na charakter sieci, jest ściśle związane z charakterem jednostek sieci. Tak więc połączenia mogą się różnić kierunkiem przepływu informacji, mogą być jednokierunkowe lub wielokierunkowe. Dodatkowo sieci złożone często ewoluują zmieniając swoją strukturę przez rozszerzanie i zanikanie połączeń. Wobec tak różnych możliwości interpretacji i złożoności tych struktur zadaje sobie sprawę ze znikomości ilości informacji jakie zgromadziłam podczas badań nad opisywanymi sieciami. Jednakże w świecie nauki nie kryje się jednak trudności, jakie stoją przed badaczami sieci złożonych. Niektóre z

nich najprawdopodobniej nigdy nie zostaną pokonane. Modele matematyczne pozwalają bowiem wyjaśniać zjawiska post factum - np. umożliwiają zrozumienie w kategoriach ilościowych awarii systemu energetycznego i pozwalają odkrywać ukryte struktury społeczne, sieci powiązań łączące świat biznesu i polityki. Rzeczywistość sieci jest jednak zbyt złożona, by można było stworzyć w pełni zdeterminowane modele zachowań. Mam jednak nadzieję, że, przedstawiony przeze mnie sposób badania pomoże w identyfikacji struktur w bardziej zaawansowanych pod względem złożoności sieci.



## Dodatek

Dodatek zawiera krótki opis technologii użytych do zbudowania programu „Analiza sieci”. Przedstawiam w nim podstawowe zagadnienia dotyczące stosowanych narzędzi programistycznych (Matlab, Perl, Macro exelowe).

Zbadanie właściwości i struktury różnego rodzaju sieci złożonych ułatwił mi zaprojektowany przeze mnie program. Przy jego budowie korzystałam z narzędzi takich jak macra w Excelu do konwersji danych tak aby ułatwić pracę programowi Grebber.pl, który z kolei służy do pobierania danych z portali (w tym przypadku nasza-klasa.pl). Efektem działania tych dwóch programów są dane przygotowane w odpowiedniej formie do pobrania przez program „Analiza sieci” wykonany za pomocą narzędzia programistycznego Matlab. Dla osób chcących zapoznać się z działaniem aplikacji proponuje zapoznanie się poniższymi podrozdziałami, w których omawiam ogólne pojęcia programistyczne, które warto poznać przed wglądem w kody programów.

### Zagadnienia dotyczące języka programowania Perl.\*

#### *Dlaczego język Perl?*

Perl jest doskonałym narzędziem do obróbki tekstu i plików binarnych. Nazwa perl jest skrótem od Practical Extraction and Report Language. Jest on bezpłatny i można go skompilować na prawie wszystkich architekturach i systemach operacyjnych. Perl jest szybszy od większości języków interpretowanych. Przed wykonaniem skryptu, program perl czyta zawartość skryptu do pamięci i „kompiluje” go do wewnętrznego formatu.

#### *Podstawowa składnia*

Pierwsza linia skryptu w Perlu zawsze wygląda następująco:

```
#!/usr/bin/perl
```

W zależności od tego gdzie znajduje się interpreter. Każda komenda musi kończyć się średnikiem (;). Tekst rozpoczynający się od znaku hash (#) jest traktowany jako komentarz. Bloki kodu, które obejmują ciało warunkowe lub pętle są ograniczane nawiasami ({...}).

---

\* Rozdział napisany na podstawie [13]

## ***Typy danych***

### **Zmienne**

Zmienna jest podstawowym typem w Perlu. Zmienna może mieć wartość całkowitą, zmiennoprzecinkową, lub znakową. Perl ustala typ zmiennej z kontekstu. Zmienne zawsze posiadają prefiks \$. Przykładowo.:

```
$str = "Hello world!";
```

W Perlu alfanumeryczny ciąg znaków bez prefiksu jest traktowany literalnie.

### **Wektory zmiennych**

Perl zawiera wektory (lub listy) zmiennych. Bieżącą wartość udostępnia prefiks @. Można również przypisać elementy wektora przez nazwę:

```
@numbers = (3,1,4,1,5,9);  
@letters = ("this","is","a","test");  
($word,$another_word) = ("one","two");
```

Można się odwoływać do poszczególnych elementów, przy czym pierwszy element ma indeks 0:

```
$blah[23] = 2.718281828;
```

### **Skojarzone wektory zmiennych**

Jest to najbardziej użyteczna cecha Perla. Można dzięki niej tworzyć tablice użytkowników według "login name" i tablice nazw plików. Prefiksem dla tego rodzaju zmiennych jest znak procenta (%). Kluczem dla tych wektorów są zmienne znakowe:

```
%quota = ("root",100000,  
          "pat",256,  
          "fiona",4000);
```

Odwoływanie się do elementów ma następującą postać:

```
$quota{dave} = 3000;    # dave - klucz, 3000 - wartość
```

## ***Najważniejsze operatory***

```
** Operator wykładniczy  
**= Przypisanie wykładnicze  
() Zerowa lista wektora  
. Połączenie dwóch ciągów znakowych  
.= Przypisanie połączenia  
eq Równość ciągów znakowych (odpowiednik ==)
```

=~ Pewne operacje na domyślnym ciągu znaków lub \$\_. Przy czym argument z prawej strony to ciąg znaków, zaś z lewej specyfikacja akcji(m.in. szukaj, wymień, ...)

- x Operator powtórzenia
- .. Operator zakresu
- f, -x, -l, ... Unarne operatory testowania pliku

## **Wyrażenia regularne**

Perl był dla mnie odpowiednim narzędziem do przeszukiwania stron internetowych ze względu na rozbudowane i nieskomplikowane funkcje wyrażen regularnych. Podczas tworzenia wyrażen pomocne okazały się informacje:

```
. dowolny znak z wyjątkiem nowej linii
c* zero lub więcej razy powtórzony znak c
c+ raz lub więcej znak c
c? zero lub raz znak c
[klasa] dowolny znak z klasy
\w znak alfanumeryczny (z "_")
\W nie alfanumeryczny (z "_")
\b granice słowa
\B bez granic
\s biały znak
\S nie biały znak
\d cyfra
\D nie cyfra
^ początek linii
$ koniec linii
| alternatywa
{n} dokładnie n razy
{n,} conajmniej n razy
{n,m} pomiędzy n a m razy
```

Przykładem komendy może być: m/wzor/cgimosx - co oznacza:

```
"c" - nie resetuj pozycji, gdy brak dopasowania, przy użyciu "g"
"g" - globalnie, tj. wszystkie wystąpienia
"i" - ignoruj wielkość liter,
"o" - wzorzec nie będzie się zmieniał podczas pracy skryptu,
"m" - dopasuj ^ i $ do każdej linii, a nie do całego wyrażenia,
"s" - potraktuj nową linię jako znak ( "." normalnie nie jej zawiera),
"x" - pozwól na wystąpienie we wzorcu spacji i komentarzy #.
```

## **Operacje na plikach**

W Perlu można korzystać z operacji na plikach tekstowych. Perl może czytać i zapisywać pliki tekstowe przez deskryptory. Są one zwyczajowo pisane dużymi literami. Pliki otwiera się komenda open. Ma ona dwa argumenty: deskryptor pliku i jego nazwę. Kolejne linie są czytane przez użycie deskryptora w ostrych nawiasach (<...>):

```
open(F, "data.txt");
while($line = <F>){
    # zrob cos z linia
}close F
```

Argument reprezentujący nazwę pliku może posiadać jeden z kilku prefiksów. Jeśli jest to <, plik jest otwarty do czytania (domyślny prefiks), > do pisania - jeśli już istnieje, jest nadpisywany, zaś >> do dopisywania.

W powyższym rozdziale opisałam tylko niektóre z wielu funkcji języka Perl, szczególnie te, które sama wykorzystałam w programie grebber.pl.

## **Krótko o programowaniu w Matlabie.**

### ***Podstawowe typy danych w Matlabie*** \*\*

Wielką zaletą Matlabu jest to, że zmiennych nie trzeba wcześniej deklorować. Matlab pozwala również na proste operowanie macierzami. Na przykład generacja wektora o elementach zmieniających się od 1 do 100 co 1 wygląda następująco:

```
>>w=[1:100]
```

Wektor o elementach od 3 do 10 zmieniających się co 0.5 jest generowany poleceniem

```
>>w=[3:0.5:10]
```

Argumenty funkcji oraz wskaźniki do elementu macierzy umieszcza się w nawiasach okrągłych, zaś elementy macierzy w nawiasach kwadratowych. Macierze i wektory możemy definiować wpisując kolejne elementy w następującej postaci:

```
>>B=[1 3 5 9]      %% (wektor 1 3 5 9)
```

Dostęp do elementów macierzy odbywa się poprzez określenie wiersza i kolumny wyrazu

```
>>B(2,3)
```

### ***Definiowanie własnych poleceń i funkcje w Matlabie*** \*\*

Bardzo wygodnym narzędziem jest możliwość zapisywania sekwencji poleceń w pliku tekstowym o rozszerzeniu .m. Reprezentuje on polecenie, które wykonuje instrukcje zawarte w pliku. W ten sposób działają standardowe komendy i funkcje programu Matlab. Przy

---

\*\*Rozdział na podstawie [15]

tworzeniu własnych poleceń lub funkcji wygodnie jest posługiwać się komentarzami. Znakiem początku komentarza jest znak %.

Wybrane przykładowe funkcje, które można wykorzystać w środowisku programistycznym Matlaba:

abs	wartość bezwzględna, moduł liczby zespolonej
ceil	zaokrąglenie w kierunku +nieskończoności
exp	funkcja eksponentialna
fix	zaokrąglenie w kierunku zera
floor	zaokrąglenie w kierunku -nieskończoności
log	logarytm naturalny
rem	reszta z dzielenia
round	zaokrąglenie
sign	znak funkcji
sqrt	pierwiastek kwadratowy
eye	tworzenie macierzy jednostkowej
ones	macierz jedynekowa
zeros	macierz zerowa
rand	macierz losowa o rozkładzie równomierny

## **Pomocne makra w Excelu i Microsoft Visual Basic**

Jeżeli jakieś zadania są często powtarzane w programie Microsoft Excel, to można zautomatyzować ich wykonywanie za pomocą makra. Makro jest serią poleceń i funkcji przechowywanych w module języka Microsoft Visual Basic i może być uruchomione zawsze, gdy trzeba wykonać dane zadania. W swojej pracy wykorzystuje markę dla ułatwienia sobie pracy przy wpisywaniu danych o połączeniach. Makro konwertuje dane wpisane w formie ciągów liczb na dwie kolumny, z których każda opisuje jedno połączenie.

## Bibliografia:

- 1) E. Benedykt (2005), *Wszystko jest siecią - Świat na sześć osób*, <http://polityka.onet.pl>
- 2) Magdalena Pecul (1998), *Jaki ten świat mały*, *Wiedza i Życie* 10/1998 , s. 13-18
- 3) I. Simonsen, K. A. Eriksen, S. Maslov, K. Sneppen (2004), *Diffusion on complex networks: a way to probe their large-scale topological structures*, [www.sciencedirect.com](http://www.sciencedirect.com), s. 163-173
- 4) Anna Radwan, *Nowe spojrzenie na topologie sieciowe*, *Nowe Technologie* (2/2005) , strony, s. 1-6.
- 5) Ł. Nowak (2007), *Analiza sieci społecznych i dynamika sieciowa*, Instytut Socjologii Uniwersytetu Warszawskiego, Warszawa
- 6) R. J. Wilson (2002), *Wprowadzenie do teorii grafów*, Wydawnictwo Naukowe PWN, Warszawa
- 7) Strona internetowa <http://wikipedia.org>
- 8) A. L. Barabási, Z. N. Oltvai (2004), *Network biology: Understanding the cell's functional organization*, *Nature reviews*, s.104-108
- 9) Katarzyna Sznajd – Weron, *W sieci małego świata*, *Wiedza i życie* (2/2004), s. 68-71
- 10) D. J. Watts, S.H. Strogatz, *Collective dynamics of 'small world' networks*, *Nature*, vol. 12, nr 393, s. 440-442
- 11) A. Fornczak, J. A. Hołys (2007), *Sieci ewoluujące: od fizyki do Internetu - Wykład 1*, <http://www.if.pw.edu.pl>
- 12) E. Miles (2007), *Yahoo hires economics, sociology researchers*, <http://news.com>
- 13) K. Juskiewicz (2006), *Kurs języka Perl*, <http://www.kt.agh.edu.pl/~juskiew/perl/>
- 14) Strona internetowa <http://www.kdm.wcss.wroc.pl/wiki/index.php/Matlab> A. Fornczak, J. A. Hołys (2007), *Sieci ewoluujące: od fizyki do Internetu*
- 15) B. Mrozek B., Z. Mrozek (2004), *Matlab i Simulink - Poradnik użytkownika*, Wydawnictwo HELION, Warszawa